NBER WORKING PAPER SERIES

THINKING VERSUS DOING: COGNITIVE CAPACITY, DECISION MAKING AND MEDICAL DIAGNOSIS

Benjamin R. Handel
Louis-Jonas Heizlsperger
Jonas Knecht
Jonathan T. Kolstad
Ulrike Malmendier
Filip Matějka

Working Paper 35034
http://www.nber.org/papers/w35034

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026

We thank David Chan and David Silver for helpful discussant comments. We thank seminar participants at the 2024 AEA meetings, the 2025 AEA meetings, Amazon, Berkeley, BYU, Duke, CESifo Summer Institute, Georgetown, Harvard, Harvard Business School, Imperial College London, LSE, MIT, Northwestern, SITE, Stanford, University of Chicago, Universidad de los Andes Bogota, University of Zurich, University of Tokyo, UCLA, and UC Riverside for their helpful feedback. Funding from the National Institute on Aging (NIA) grant P01AG005842 to the National Bureau of Economic Research and from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant No 101002898) are gratefully acknowledged. We thank Julia Adler Milstein for her collaboration in accessing the data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Thinking versus Doing: Cognitive Capacity, Decision Making and Medical Diagnosis
Benjamin R. Handel, Louis-Jonas Heizlsperger, Jonas Knecht, Jonathan T. Kolstad, Ulrike Malmendier, and Filip Matějka
NBER Working Paper No. 35034
April 2026
JEL No. D83, D91, I11

ABSTRACT

We study how situational fluctuations in cognitive capacity shape behavior in high-stakes, real-time decision-making. Drawing on recent advances in behavioral economics that revolve around inattention, cognition and complexity, we show that cognitive load influences how physicians in emergency departments allocate mental effort and attention when making diagnostic and treatment decisions. We use quasi-random variation in patient-physician pairings, along with granular electronic medical record and audit-log data from many clinical interactions, to show that, under higher cognitive load, physicians substitute mental deliberation with more numerous but less precise diagnostic actions. Specifically, we document that higher load (i) increases the total number of orders of diagnostic tests (ii) reduces the use of targeted, but more uncommon tests (iii) increases the use of common tests and (iv) increases uncertainty in diagnostic beliefs. Cognitive load impacts downstream inpatient admission from the emergency department: a physician in the highest cognitive load decile increases admissions by 28% relative to the same physician in the lowest cognitive load decile, for the exact same kind of patient. These results offer novel field-based evidence on the dynamics of attention and belief formation, and shed light on how cognitive constraints shape diagnostic behavior in complex, real-world environments.

Benjamin R. Handel
University of California, Berkeley
Department of Economics
and NBER
handel@berkeley.edu

Louis-Jonas Heizlsperger
University of California, Berkeley
louis.heizlsperger@berkeley.edu

Jonas Knecht
University of California, Berkeley
jonasknecht@berkeley.edu

Jonathan T. Kolstad
University of California, Berkeley
Haas School of Business
and NBER
jkolstad@berkeley.edu

Ulrike Malmendier
University of California, Berkeley
Department of Economics
and NBER
ulrike@econ.berkeley.edu

Filip Matějka
CERGE-EI
Filip.Matejka@cerge-ei.cz

1 Introduction

All decision-making requires mental effort. Agents in the changing and complex economy need to filter out new information from numerous sources (e.g., media, customers, coworkers, data) and integrate it with their existing knowledge to form their behaviors (e.g., on the job decisions, personal choices, tacit learning). These evaluations and decisions often occur under conditions with significant bandwidth constraints, as well meaningful levels of institutional complexity. Recent advances, mostly theoretical and experimental, highlight the importance of inattention, cognitive capacity, and complexity in shaping economic decisions (e.g., Maćkowiak et al. (2023), Gabaix (2019), Bordalo et al. (2025), Caplin and Dean (2015), Oprea (2024) and Enke and Graeber (2023)). In fact, integrating such cognitive factors into economic models can potentially unify seemingly disparate behavioral phenomena under a common framework.

A key gap in the literature, however, is the lack of empirical exploration of high-frequency decision-making in real-world settings. Such empirical investigations are not only valuable for their applied relevance, but also help to inform and refine existing models of cognitive resource allocation. In this paper, we begin to fill this gap by exploring a practical, high-stakes, high-frequency setting: the diagnostic choices of physicians in emergency departments (EDs). In this setting, physicians must obtain information from multiple sources (e.g., patient history, physical examinations, conversations, imaging, and lab tests), synthesize these data into diagnostic hypotheses, and ultimately arrive at a treatment plan.

These decisions are cognitively demanding and made under time pressure, uncertainty, and institutional complexity.[1] Importantly, physicians treat hundreds or thousands of patients over time, many of whom present with similar symptoms, offering an opportunity to study decision variation within-physician over repeated, comparable choice settings. Since cognitive capacity is difficult to measure in administrative data, much of the prior literature on cognitive load has focused on controlled laboratory settings, where cognitive load or task complexity can be systematically manipulated. However, medical diagnosis presents a unique case: the cognitive effort involved in forming a diagnosis ("thinking") is often accompanied with reliance on diagnostic tests ("doing"). Because the use of diagnostic tests is observable, it provides a valuable empirical window into the underlying cognitive mechanisms that precede testing.

We show that fluctuations in cognitive resources meaningfully affect physicians' behavior over time, even for the same physician facing similar decisions. Variation in cognitive effort influences key aspects of physicians' clinical practice, including the number and precision of diagnostic test orders, as well as costly inpatient hospital admission decisions. Beyond providing insight into the

[1] Since the seminal work of Arrow (1963) a large literature has developed to study aspects of physician decision-making (see, e.g., McGuire (2000) and Chandra et al. (2011) for overviews). This work is largely grounded in standard economic models of incentives, information, and market structure with little to no focus on within-physician variation in behaviors and, especially, the cognitive constraints of physicians.

diagnostic process in the ED, our findings yield broader behavioral conclusions: the cost of mental effort for experts is not fixed over time, and care approaches change flexibly, even over short time horizons. Moreover, physician behavior exhibits attenuation—a central feature of inattention-based models—where diagnostic responses under high cognitive load are systematically muted relative to those under lower load.

Specifically, we (i) develop a conceptual framework that formalizes the role of cognitive resources in diagnostic decision-making and (ii) leverage new granular data from the Emergency Department of the University of California at San Francisco (UCSF) to empirically study the implications of cognitive capacity on care. Our conceptual framework models an individual physician's pursuit of a patient's diagnosis. The physician has a level of cognitive load throughout the patient encounter that depends on the other patients she is treating or has treated. The patient arrives with a chief medical complaint and some other characteristics, from which the physician forms a prior about potential diagnoses. In sequence, the physician (i) exerts mental effort to refine her diagnostic beliefs and (ii) orders external tests to further refine those beliefs. The marginal costs of mental effort by the physician are increasing in both cognitive load and current mental effort. The physician exerts mental effort and orders external tests to the point where the benefits of further refining beliefs about the true diagnosis are outweighed by the costs of additional mental effort and external tests. Finally, the physician makes an inpatient admission decision, trading off the high cost of admission with the incremental benefit of ensuring appropriate treatment. Some of our modeling choices reflect the specifics of our empirical environment and we present an extended discussion of how our model relates to prior conceptual work on cognitive load.

The model has several predictions that we directly test in our empirical analysis. First, the model predicts that a higher level of cognitive load leads to the physician exerting less effort to acquire mental signals and placing more external orders instead. Second, the model predicts that higher cognitive load reduces the use of more infrequent, targeted tests and increases the use of more frequent, broad tests, reflecting behavioral attenuation. Third, the model predicts that cognitive load increases uncertainty in diagnostic beliefs, both before and after choosing external diagnostic tests. Fourth, the model predicts that higher cognitive load increases inpatient hospital admissions from the emergency department, a costly and consequential treatment action. For each of these model predictions, we develop and test an empirical analog in our data.

The data include rich information on both patients and providers including (i) patient characteristics, (ii) medical care (e.g., orders, procedures, triage, diagnoses), (iii) physician characteristics, (iv) note actions where physicians take clinical notes on patient encounters and (v) an audit log of physician click-by-click actions when placing medical orders / notes or reviewing the outcomes of prior orders / notes. These data include all patients visiting the Emergency Department over a two-year period (mid-2017 to mid-2019) and include to-the-second timestamps for all relevant actions. Crucially, the detailed data allow for the construction of short-duration within-physician-shift cognitive capacity measures while our environment in the ED generates quasi-random variation in those measures over time within-physician.

Our dataset has several features that make it very well suited for the exploration and application of theories of cognitive resource allocation. First, physicians repeatedly face distinct diagnostic decisions across thousands of patient encounters, allowing us to study joint statistical properties of patients' states and physicians' choices. Second, the decision space is rich, with many possible diagnoses, diagnostic orders, and treatment options. This richness is not only important for studying flexibility in allocation of cognitive resources but it is also essential to be able to address "thinking" and infer physicians' beliefs, which are revealed by the detailed choices they make during the diagnostic process. Third, the environment involves trained professionals, enabling us to focus on cognitive limits of short-term decisions rather than on humans' long-term learning about how to approach a particular task at hand.

Empirically, we resolve two key identification issues in order to establish the impact of cognitive load on different outcomes. First, we address the issue that actual observed outcomes for a given patient are endogenous to the level of their physician's cognitive load. We develop a model of predicted patient complexity / intensity based on granular information ex ante to the patient being treated by the physician. This information includes details like (i) patient chief complaint at a granular level, (ii) patient vital signs, and (iii) the triage acuity score given to the patient by the intake nurse. We use a LASSO model that predicts, given this information for each patient, the expected (i) number of orders, (ii) amount of provider time editing notes, and (iii) number of outside consults to specialists. The output of this prediction problem is a measure of expected patient complexity / intensity that does not depend on the patient's actual experienced treatment.

For a given provider, we then aggregate these patient-level predictions within-shift and across patients into granular measures of cognitive load. Our primary measure of cognitive load relates to the recent burden of complex patients a physician treats. For a given time window (e.g., 90 minutes), we assess the total complexity burden for a physician by summing up predicted patient complexity across all patients the physician is currently treating (excluding the focal patient). For a given physician and given shift, at each point in time we have a measure of contemporaneous burden as well as lagged burden from earlier in the shift. We utilize both the current and lagged measures in our empirical analysis to reflect physician cognitive load. For each patient-physician pair, we construct these measures "leaving out" the contribution of the index patient we are studying. We assess robustness for these measures on a number of key dimensions including, e.g., (i) the time intervals used (ii) the features that enter our patient intensity prediction model and (iii) variations of (and alternatives to) the underlying LASSO prediction framework.

With these measures in hand, the second identification challenge we resolve is ensuring that we have quasi-random variation in cognitive load, within and across physicians. To this end, we leverage both quasi-random patient arrival to the ED as well as quasi-random assignment of patients to physicians once they have arrived in the ED.[2] We present evidence on within-physician variation in cognitive load over the two years to illustrate the significant variation coming from quasi-random patient arrival. We present balance tests on patient characteristics to highlight the quasi-random nature of patient assignment to physicians, which we also support with qualitative evidence on the assignment process.

We set up our main analyses at the provider-patient-event level. We study the impact of provider cognitive load on key outcomes, conditioning on time of day, time of shift, day of week, acuity code and chief complaint fixed effects. In addition, we run most specifications with provider fixed effects, isolating the impact of cognitive load within-provider over time, relative to their baseline behaviors. A critical aspect of our model is the nature of time in the treatment process: we focus on current cognitive load by measuring the level of expected complexity for assigned patients in the prior 90 minutes when an action (e.g. ordering a test) is made. We include lags related to cognitive load prior to that period to control flexibly for the stock of cognitive load.[3] We allow for the effects of cognitive load to enter linearly as well as more flexible indicators for quintiles and deciles.

We first test our model's prediction that higher cognitive load will lead to an increase in the number of diagnostic tests. Our main outcome variables are the number and types of tests in the order batches for each patient.[4] We find that increasing physicians' contemporaneous cognitive load by 1 standard deviation has a meaningful effect on the number of diagnostic tests in an order batch, increasing them by 4% holding all else fixed. As a placebo, we estimate the same model for medication orders, which are more standardized and require less thinking, finding a precise null effect of cognitive load on these orders. For our specification using quintiles of physician cognitive load, we find that the top quintile of cognitive load increases the total number of diagnostic orders by 9% relative to the bottom quintile.

We then test our model's second prediction, digging into how the profile of diagnostic orders changes with respect to cognitive load. We find that the overall increase in diagnostic orders due to higher cognitive load masks a significant underlying shift from less common orders to more common orders. A 1 standard deviation increase in cognitive load increases the number of common orders (top 25 diagnostic orders) by 7% but decreases rare orders (outside of top 25) by 5%. In our quintiles specification, we find that moving from the bottom to the top quintile of cognitive load increases common orders by 19% and decreases rare orders by 16%.

Next, we test our model's third prediction, that diagnostic uncertainty increases under higher

[2]Our analysis here follows a range of recent papers that use quasi-random variation in ED patient arrival and assignment to study different questions, including, e.g., Chan (2016), Chan (2018), Mullainathan and Obermeyer (2019), and Silver (2020).

[3]We show in Appendix F.4 that our primary results are robust to alternative time windows including 60 and 120 minutes.

[4]Medical orders are typically placed in batches, where a number of orders are input simultaneously. We use this as a natural unit of observation, unpacking the characteristics of these order batches as described.

cognitive load. Building on our model and work by Caplin and Dean (2015), we develop a novel empirical approach to study diagnostic belief refinement. We base our approach on statistical entropy reduction, a measure used in machine learning and information theory that quantifies the level of information gain via uncertainty reduction. We do this for a given set of diagnostic orders, assessing the extent to which those orders move us from an ex ante prior towards a final diagnosis. This in turn tells us about the extent of information the physician has incorporated since being assigned to the patient.

We leverage our detailed data to implement this technique. We proceed in two steps. First, for each possible order and ex ante information set (e.g., chief complaint) we construct the probability distribution of final diagnoses conditional on placing that order. For an order batch, we combine these distributions across the orders in the batch. Second, we compute the Shannon entropy of this implied final diagnosis distribution, which gives us a measure of how targeted the order batch is towards uncovering specific final diagnoses. Our key measure focuses on the entropy reduction of the first order batch. This measures, relative to the distribution of possible diagnoses associated with the ex ante information set (chief complaint and acuity assessment), how much more targeted is the first order batch towards particular diagnoses. The higher the entropy (lower reduction in uncertainty from the prior) the less precise physician's beliefs are and the less the physician learned about the diagnosis up to the moment of placing orders.

We find that a 1 standard deviation increase in cognitive load decreases the entropy reduction of the first order batch by 2%. In the specification with quintiles of cognitive load, being in the top quintile of cognitive load leads to a 5% decrease in entropy reduction relative to the bottom quintile of cognitive load. Much of this change occurs moving from the fourth to the fifth (top) quintile of cognitive load, illustrating the non-linear relationship between our cognitive load measure and the targeting of diagnostic tests.

We next test our model's fourth hypothesis: that higher cognitive load will lead to more inpatient admissions. To do this, we run regressions at the provider-patient encounter level, since each outcome relates to the whole course of treatment in the ED. We use the maximum point-in-time cognitive load for the physician over the entire patient encounter as our measure of cognitive load.

We find a large effect of increased cognitive load on hospital admission: under our linear model a 1 standard deviation increase in maximum cognitive load increases the chance of admission by 9%, controlling for provider, chief complaint, acuity level, day and time fixed effects. In our more flexible specification with cognitive load deciles, we find a large and near monotonic effect of cognitive load moving from low to high deciles. The admission probability of a patient is 28% higher for a physician in the highest cognitive load decile relative to when in the lowest decile.

We assess the mechanisms underlying this result and find that this large impact of cognitive load is not due to ED capacity constraints being correlated with situations of higher cognitive load.

Our results on reduced diagnostic precision under higher cognitive load suggest that provider risk aversion is a plausible explanation for these admission increases under higher cognitive load. We perform two subsequent analyses to support this. First, we follow Abadie (2003) and characterize the marginal patients admitted. Across a range of metrics, this analysis shows that the marginally admitted patients due to higher physician cognitive load are healthier than the patients typically admitted. Second, we study the impact of cognitive load on hospital readmissions and find a precise zero effect for the impact of maximum cognitive load on hospital readmissions within 30 days. Both of these analyses are consistent with the hypothesis of increased provider risk aversion leading to more hospital admissions when providers have higher cognitive load.

We perform a range of additional analyses. We run our primary analyses for two specific chief complaints: (i) abdominal pain (9% of our sample) and (ii) chest pain (4% of our sample). These are two of the most common chief complaints and also have meaningful degrees of freedom in diagnosis and treatment. These results corroborate our primary findings and allow us to investigate diagnostic impacts with additional contextual specificity. We also run encounter-level regressions for all key outcomes that use maximum cognitive load over the encounter, with results that support (and often exceed) our primary event-action-level results. For example, the encounter-level analysis using deciles of cognitive load shows a strong and monotonic relationship on the impact of cognitive load on the number of diagnostic orders, with the top decile placing 42% more diagnostic orders than the bottom decile. We also investigate an additional outcome, note edit time. We find that a 1 standard deviation increase in cognitive load decreases note edit time for the index patient by 7%, and that moving from the bottom to top quintile of cognitive load reduces that note edit time by 23%. Additionally, we assess the heterogeneous impacts of cognitive load using a causal forests approach following Wager and Athey (2018) and Athey et al. (2019). While the average treatment effects replicate our main findings, the heterogeneous effects are generally underpowered and don't paint a clear picture of differences in cognitive load impacts across providers. Finally, we perform a counterfactual calculation where we ask how a scheduling algorithm that accounts for cognitive load impacts the diagnostic tests ordered as well as the rate of hospital admissions. Holding fixed the set of physicians in the ED at a point in time as well as the number of patients assigned to each, we show that this kind of reallocation has a large impact on tails of high cognitive load, where treatment effects are strongest. This leads to increased order precision, decreased number of orders, and reduced hospital admissions, without meaningfully changing the type of underlying physician treating patients.

This paper contributes to the growing theoretical and empirical literatures on attention, cognitive load and complexity in decision making. Maćkowiak et al. (2023) presents an overview of rational inattention and how it relates to phenomena studied in behavioral economics, highlighting the conceptual implications for decision-making and empirical evidence in this area. Gabaix (2019) also provides a conceptual overview in the space of behavioral inattention. Enke and Graeber (2023) presents an overview of related work in this area focusing specifically on complexity and cognitive burden, highlighting advances and recent empirical work, the vast majority of which is in a lab experiment context including, e.g., Dean and Neligh (2023), Oprea (2024) and Xiang et al. (2021). Bordalo et al. (2025) present a broader framework for cognitive load where individuals (i) categorize current problems based on past experiences and (ii) make decisions based on weights determined by that categorization. In Section 2 we compare our framework to these prior papers, noting ways in which their frameworks could generate additional testable hypotheses in future work. Relative to this literature, a key contribution of our framework is to develop a conceptual cognitive load model that maps to the provider diagnostics in the ED and then to directly test model hypotheses using detailed administrative data in a high-stakes setting.[5]

While there are many health care papers focused on cross-sectional variation in provider behaviors (e.g., due to skill, preference, or training differences) there are fewer papers that study within-physician behavior changes, and very few that study physician cognitive load in depth.[6] Chan (2018) and Silver (2020) study within-physician behavior changes in the ED context. Chan (2018) finds that physicians near the end of their shifts slack off by taking on fewer patients and spending more per patient while Silver (2020) finds that physicians change their behaviors in response to their peer groups and assimilate towards group production norms. Shanmugam (2020) studies the impact of physicians' cognitive capacities on health care equity in New York state EDs using variation similar to what we leverage here. She finds that, when cognitively constrained providers shift within-ED treatments and admissions towards high-risk, uninsured patients but shift diagnostic testing in the opposite direction. Her context has greater external validity (patients in an entire state) but data that are less deep, leading to a coarser analysis of cognitive load and its impacts. Agarwal et al. (2023) perform an experiment assessing radiologist practices change in response to information interventions, focusing on the interplay between human expertise and advice from artificial intelligence (AI) tools. Other notable papers studying within-physician behavior changes include Steiny Wellsjo (2025) and Chodick et al. (2025).

The remainder of the paper proceeds as follows. Section 2 presents our conceptual framework, studying the impact of cognitive load on key behaviors and outcomes. Section 3 describes our data and context. Section 4 sets up our main analysis by constructing our cognitive load measures and assessing threats to identification. Section 5 presents our main empirical specification. Section 6 presents our results and and Section 7 concludes.

[5]There are papers in other domains that use proxies for cognitive load to study the impacts on expert behavior including, e.g., for judges (Danziger et al. (2011), Kleinberg et al. (2017)) and manufacturing workers (Kaur et al. (2025)).

[6]See Skinner (2011) and Cutler et al. (2019) for surveys of the physician practice variations literature. See, e.g., Chandra and Staiger (2007), Finkelstein et al. (2016), and Badinski et al. (2023) for specific examples. Several recent papers hone in on the diagnostic process in a cross-sectional sense, including, e.g., Song et al. (2010), Abaluck et al. (2016), Currie and MacLeod (2017) Mullainathan and Obermeyer (2019), and Chan et al. (2022). These papers illustrate meaningful cross-sectional variation in diagnostic practices and unpack different underlying mechanisms. Doyle et al. (2010) is notable in showing that patients who are randomized to physicians from higher-ranked institutions have lower costs, primarily due to lower diagnostic testing rates, while achieving similar outcomes.

2 Conceptual Framework

In this section, we develop a model of an ED physician who resolves diagnostic uncertainty and chooses treatment. We then derive the model's implications, connect its elements to observables in the data, and use the model to interpret our empirical findings. Finally, we state testable hypotheses and discuss the model's simplifying assumptions.

2.1 Model

A physician faces a patient of an unknown state (diagnosis) x. She chooses: (i) how to learn about state $x \in [0,1]$, (ii) what treatment $y \in [0,1]$ in ED to prescribe, and (iii) the level of admission $A \in [0,1]$ to inpatient care. She maximizes expectation of the following objective:

$$-(1-A)\big|y-x\big| - A^2 - \text{ cost of information.} \tag{1}$$

The first term represents health and resource loss from misalignment between the chosen ED treatment y and the patient's true state x. Better information allows the physician to match the optimal treatment $y^* = x$ more closely. The scalar $(1-A)$ models admission to hospital as a substitute to an appropriate treatment in ED. Admission mitigates losses from mistreatment, but it is costly - the cost is A^2, which is the second term in (1). Information costs appear in the third term and are specified below.

Information acquisition The physician first acquires mental signals from thinking about the patient and then external signals from doing diagnostic tests. She chooses how many mental signals, n^M, and how many external signals, n^E, to acquire in order to refine belief about x. The total number of signals is

$$n = n^M + n^E.$$

Let f be a pdf that denotes the physician's subjective belief about a patient's state. Let f^0 denote a prior belief, and f^k be a belief after k signals are received. We assume that f^0 is uniform on $[0,1]$. We assume that all signals have the form of a binary partition. Given a current belief f^k is $U[a,b]$, the physician finds out whether the state is in the lower or the upper half of the domain, i. e., f^{k+1} is either $U[a, \frac{a+b}{2}]$ or $U[\frac{a+b}{2}, b]$. After n signals, posterior f^n is uniform over an interval of length 2^{-n}.

Cost of information. The physician faces the following cost of information:

$$C(n^M, n^E, L) = C^M(n^M, L) + C^E(n^E). \tag{2}$$

C^M is the cost of thinking, while C^E is the cost of external signals from diagnostic orders - both are convex in the number of signals, i.e., marginal cost of each type of signals is increasing.[7]

L denotes prior cognitive load. Higher mental load of other activities reduces the physician's cognitive resources available for the current patient. We assume that marginal costs of thinking are increasing in L for all n^M. Cost of diagnostic tests are unaffected by L.

Actions and objective After receiving all the signals, the physician chooses treatment y to minimize the expectation of loss $|y-x|$ in the objective (1). Given her posterior belief f^n, she optimally sets $y = E[x]$, and expected $|y-x|$ equals $\frac{1}{4}2^{-n} = 2^{-n-2}$. Note that expectation of $|E[x]-x|$ equals $\frac{1}{4}$ of the size of a uniform belief's support. Given n, the optimal admission level is $A = 2^{-n-3}$.[8] The expected health loss is then equal to $(1-2^{-n-3})2^{-n-2}$ and the cost of admission is 2^{-2n-6}.

The physician therefore solves

$$\max_{n^M, n^E \in \mathbb{N}_0} \quad -(1-2^{-n-3})2^{-n-2} - 2^{-2n-6} - C(n^M, n^E, L). \tag{3}$$

Summary of timing:

1. Patient of an unknown state x checks in; physician holds prior f^0.
2. Physician chooses how much to think about the patient, n^M, updating beliefs to f^{n^M}
3. Physician chooses how many diagnostic orders to place, n^E, updating to f^n.
4. Physician chooses treatment $y = E[x|f^n]$ and admission $A = 2^{-n-3}$.

2.2 Effects of Cognitive Load

We now study how prior cognitive load affects the physician's behavior. All monotonicities referred to in this section are weak ones due to the discrete nature of choices of n^M and n^E.[9] Proofs are in Appendix A.

Proposition 1. ***(Thinking - doing - admissions)*** *If prior cognitive load L increases, then the physician chooses:*
(i) a smaller total number of signals n,
(ii) a smaller number of mental signals n^M,
(iii) a higher number of external signals n^E,
(iv) a higher level of admissions A.

If cognitive load increases, then the physician acquires less total information, thinks less (lower n^M), does more (higher n^E), and admits to inpatient care more frequently. Higher load increases

[7] Both types of costs being convex reflects the idea that cognitive load increases the time cost of activities as well as the ability to handle complexity.
[8] Simply by maximizing $-(1-A)2^{-n-2} - A^2$
[9] Strong monotonicities would hold if uncertainty were resolved on a continuous scale in a slightly modified model.

marginal cost of thinking, reducing both n^M and total information n. External tests partially substitute for thinking but do not fully offset it due to convexity of their costs. Finally, admission increases because it is increasing in posterior uncertainty in f^n. The more uncertain about x the physician is, the more appealing the prospect of admission becomes, because it scales down losses from mistreatment $y \neq x$.

Proposition 1 has immediate implications regarding levels of uncertainty about x and also what types of diagnostic tests the physician uses. The first is implied by the fact that since higher L decreases both n and n^M, then external signals are used at higher levels of uncertainty, i.e., when the physician knows less than if L were lower.[10]

Corollary 1. ***(Uncertainty at diagnostic orders)*** *If prior load L increases, then*
(i) uncertainty in f^{n^M} at the time of the first diagnostic order increases,
(ii) average uncertainty across the times of all diagnostic order increases.

The second immediate implication of Proposition 1 is concerned with the types of diagnostic tests. Each test is a particular type of refinement of the state-space of x.

Less frequent tests correspond to later refinement in the diagnostic sequence, resolve smaller state space of x, and are therefore ordered with lower probability. For instance, the first external diagnostic test is used after n^M mental signals and is conditioned on one of 2^{n^M} possible forms of posterior beliefs. Each one is realized with the prior probability 2^{-n^M} and leads to a different type of diagnostic tests. Lower mental effort leads to coarser targeting and higher frequency of common tests.

Corollary 2. ***(Frequency of types of diagnostic tests)*** *If L increases, the physician uses relatively infrequent tests even less often, and frequent tests more often.*

These results mean that cognitively loaded physicians tend to use diagnostic tests that are more of a "gold standard" for a general population of patients rather than for the patient at hand. Cognitive load hinders targeting of care and increases regularization. Physicians represent patients more coarsely, draw less on their mental abilities, and rely more on diagnostic tests.

2.3 Implications for Empirical Work

The key elements of our model relate to our data as follow. We observe realized diagnosis (x), treatment decisions (y), admission decision (A), and diagnostic orders, and therefore their number (n^E), too We also observe measures of current and prior cognitive load, which capture different elements of L, as described in Section 4.3.

While mental signals are unobserved, the model provides guidance on how to infer beliefs and changes in them due to the unobserved mental signals. We assume that the physician has

[10] All beliefs are uniform over an interval and thus such uncertainty can for instance be represented by the interval's size.

rational expectations. Then, the prior f^0 coincides with the empirical distribution of diagnoses (conditional on observables) [11]

The form of intermediate beliefs can be identified from the choice data. A similar approach as to identifying f^0 applies to f^k at any moment when a physician orders a diagnostic test, which we do observe. Consider an empirical distribution $g(x|test)$, where *test* is a particular type of a diagnostic test. It is a distribution of diagnosis conditional on *test* being used during an encounter. If a physician has rational expectations, then this distribution coincides with her expected beliefs across all moments when she orders this particular test, before she acquires additional information. In our model, the set of possible diagnostic tests is rich such that each test is ordered for one particular type of belief only. Therefore, beliefs of a physician with rational expectations at the moment of ordering the test are given by the empirical distribution $g(x|test)$.[12]

Lemma 1. ***(Inference of beliefs)*** *Belief f at a moment of ordering a particular diagnostic* test *coincides with an empirical distribution $g(x|test)$.*

Our model and its connection to data thus allow us to state the following hypotheses, which we can test in the empirical part of this paper:

Hypothesis 1: Cognitive load increases the total number of orders of diagnostic tests.

Hypothesis 2: Cognitive load reduces the use of infrequent tests, increases the use of frequent ones.

Hypothesis 3: Cognitive load increases uncertainty in diagnostic beliefs.

Hypothesis 4: Cognitive load increases admissions.

2.4 Discussion of Assumptions

The model is intentionally stylized to highlight mechanisms and generate testable predictions. Below we discuss how its simplifying assumptions affect the results.

Model Timing. The timing of our model (ex ante information produced, provider diagnostic thinking, provider diagnostic testing, treatment / admission) is supported by research on emergency department production. Pelaccia et al. (2014) discuss how physician evaluation and hypothesizing plays a crucial role in the diagnostic process, in between receiving intake information and ordering diagnostic tests. They also highlight the important role of physician cognitive processing

[11] In the empirical analysis, this is conditioned on observables such as chief complaint.

[12] This is analogous to a revealed preference approach in Caplin and Dean (2015).

at this phase. This is supported by myriad other papers (e.g., Paley et al. (2011)), UCSF's patient facing documentation (UCSF Emergency Department, 2026) and the websites of many EDs.

Form of information. The model assumes a very particular structure of information acquisition based on binary partitions. However, results in Proposition 1 or its corollaries do not depend on this stylized specification. We could use information structures that deliver richer sets of posteriors via more general Blackwell experiments. In our model, cognitively loaded physicians represent patients more coarsely. More general experiments could introduce noise of various forms, e.g., patients could also be misrepresented. The main driver of the results is that different sources of information, i. e., mental and external, are partial substitutes and have different cost sensitivities to cognitive load. The results would break if all mental as opposed to external signals resolved completely different aspects of uncertainty about x and thus the two sources of information were not substitutes at all. Additionally, the results would be different in states of the world where mental signals are complements to external signals, i.e., cases where thinking more up front enhances the effectiveness of specific external tests for facilitating treatment.

Cost of information. Our results rely on the assumption that the marginal cost of additional mental signals is increasing in L, while cost of external diagnostic tests is unaffected by it. It seems intuitive that thinking is more affected by prior load than ordering laboratory and other diagnostic tests. Let us emphasize, that in our framework mental signals are any form of information acquisition that we do not observe in data on diagnostic tests. It can be innovative deliberation, drawing on memory or experience, or even asking a colleague.

Richness of tests. Lemma 1 on inference of beliefs rests on the assumption that there are many types of tests that are available and useful under different diagnostic beliefs. If, instead, the same test were ordered across many beliefs, then an empirical distribution of x conditional on the test would only identify *average* beliefs across these situations weighed by their relative frequency. [13] While full richness is unlikely in practice, our data contain an unusually large set of diagnostic tests, allowing detailed differentiation across belief states. We discuss this in more depth in our empirical model of statistical entropy in Section 6.1.

Admission decision. We model admission as a mitigation device against mistreatment. A richer model could introduce heterogeneity across diagnoses x, with some conditions requiring admission as part of treatment and involving higher health stakes. Our framework is a reduced

[13]For instance, the literature on rational inattention assumes full richness, i.e., agents can choose any signals they wish, and thus beliefs can be fully identified from actions. There, a fairly general result is that each action in fact optimally corresponds just to one form of realized posterior knowledge. Actions thus identify knowledge, which can be inferred from the joint distribution of states and actions (Matějka and McKay, 2015; Caplin and Dean, 2015).

form of such environments: in both cases, admission functions as insurance against diagnostic uncertainty.

Homogeneous stakes and a continuous admission choice can be interpreted as a representative-agent approximation of a model with heterogeneous stakes across diagnoses and a binary admission decision. The qualitative implications of changes in cognitive load L would be analogous in such a model. The main difference is that information acquisition would become asymmetric across diagnostic paths, as early signals would differentially affect the value of further information.

Finally, we assume a quadratic admission cost to ensure an interior solution; other concave cost functions would yield similar results.

Effect of cognitive load. Our model relates closely to prior work on information theory and rational inattention. Cognitive load enters as an increase in the marginal cost of information.[14] Adjacent to our framework, cognitive load could alternatively affect physicians' prior knowledge, risk preferences, or the level of stakes. None of these mechanisms generate the substitution from thinking to doing predicted by Proposition 1. Higher prior uncertainty, risk aversion or stakes would intensify both mental as well external information acquisition.[15] While these factors could be relevant, empirical tests that yield results in line with the predictions from Proposition 1 are sufficient to identify the directional tradeoffs between thinking and doing we establish, subject to the effects of these other potential factors pushing in the opposite directions.

More broadly, it is useful to consider our framework in the context of other recent models related to cognition and cognitive load. Bordalo et al. (2025) present a general framework where decision-making is driven by how individuals categorize current problems based on past experiences. This process determines which features of an option receive attention and which are neglected, directly influencing valuation and choice. Applied to our setting, this could be used to test hypotheses related to how specific patient features or specific contextual features impact provider behaviors. For cognitive load specifically, one could extend the Bordalo et al. (2025) framework with, e.g., differential categorization costs, cognitive load conditional on categorization, or categorization that reflects the state of cognitive load. This could generate a range of interesting testable hypotheses, likely with high demands on the data.[16]

[14]Combined with the structure we assume on how effort impacts diagnostic precision, when someone is cognitively loaded an individual has a coarser representation the space of possible outcomes and chooses tests that are more central in the space. This produces a form of regularization where you move towards the middle, i.e. the most common tests, because the provider hasn't shrunk their prior enough to warrant testing in more targeted spaces. This implication relates to work on Bayesian regularization, e.g. in Xiang et al. (2021) and Enke and Graeber (2023), if cognitive load makes it more likely that a provider has a coarser subjective prior on a patient's diagnosis.

[15]Another possible factor we assume away is the impact of cognitive load on how the physician is able to incorporate external test results into their diagnostic process. Under different specifications, this could naturally push the provider either towards or away from more targeted tests.

[16]Possible testable hypotheses include those (i) related to heterogeneous impacts of cognitive load as shaped by provider past experiences and (ii) choice instability as a result of otherwise irrelevant cognitive frames. Bordalo et al.

In our framework, cognitive load operates by leading doctors to more coarsely represent patient/problem/symptom characteristics but does not let physicians (i) misrepresent (instead of coarsely) represent the problem or (ii) draw differently on knowledge and experience given the representation, apart from the inclusion of a physician type. We view these latter two potential impacts of cognitive load as promising areas to focus on in future work, especially in contexts, e.g., with field experiments, where it is straightforward to identify these phenomena.

3 Institutional Setting and Data

We study physician decisions for the universe of patients over age 18 who arrived at the University of California, San Francisco (UCSF) emergency department (ED) between July 1, 2017 and June 30, 2019.[17] Our primary dataset comprises 76,001 visits to the ED, which we refer to as *encounters*.[18][19] Figure 1 illustrates a simplified version of the ED workflow from the patient perspective. Appendix B describes this workflow in detail. We describe different parts of this workflow, when relevant to our analysis, as we proceed through the paper.

Upon arrival (via ambulance or walk-in), the typical patient flow has two main steps prior to being assigned to a physician. First, a registrar collects patient demographic information, solicits information on why the patient has come in, and opens the encounter in Epic. Second, a licensed triage nurse performs a focused assessment in a dedicated triage bay (or at the bedside for direct-to-room arrivals). Vital signs, symptoms, and a brief history are obtained and recorded, together with any required isolation flags. Using the Emergency Severity Index (ESI, version 5) algorithm the nurse assigns an acuity code from 1 (immediate life-saving intervention) to 5 (no resources anticipated). These scores are subsequently used for triage and physician assignment, as discussed in the next subsection. The triage nurse also selects a structured chief-complaint label from Epic's menu (e.g., "chest pain", "abdominal pain") or enters free text. There are 732 possible chief complaints at the most granular level of Epic's hierarchy, which roll up into 72 categories of chief complaints in an intermediate tier and 15 categories at the most aggregated tier.

Table 1 provides summary statistics on patient demographics, health status, and chief complaints. 50.6% of patients are male, 46% identify as white, 15% as black, and 18% as Asian. The mean (median) age is 51.3 years old (52) with a distribution that is left-skewed and covers patients from 18 to the top-coded value of 89. 31% of patients arrive via an emergency vehicle with 69% walking in (or driving in) on their own.

(2020) study a framework related to Bordalo et al. (2025), where there is history-dependent behavior that, together with a notion of varying cognitive costs, would generate an additional range of testable hypotheses in our setting.

[17] Due to privacy restrictions, we don't study minors. Adults comprise the vast majority of ED patients in our context.

[18] 38.19% of encounters involve patients with multiple visits during our observation period, yielding 46,976 unique *patients* in our sample. For simplicity, we refer to encounters and patients interchangeably.

[19] Our data overall have 84,214 encounters but 8,213 are dropped from our primary sample because they have no orders or notes. These patients are either sent home at triage or admitted to inpatient immediately.

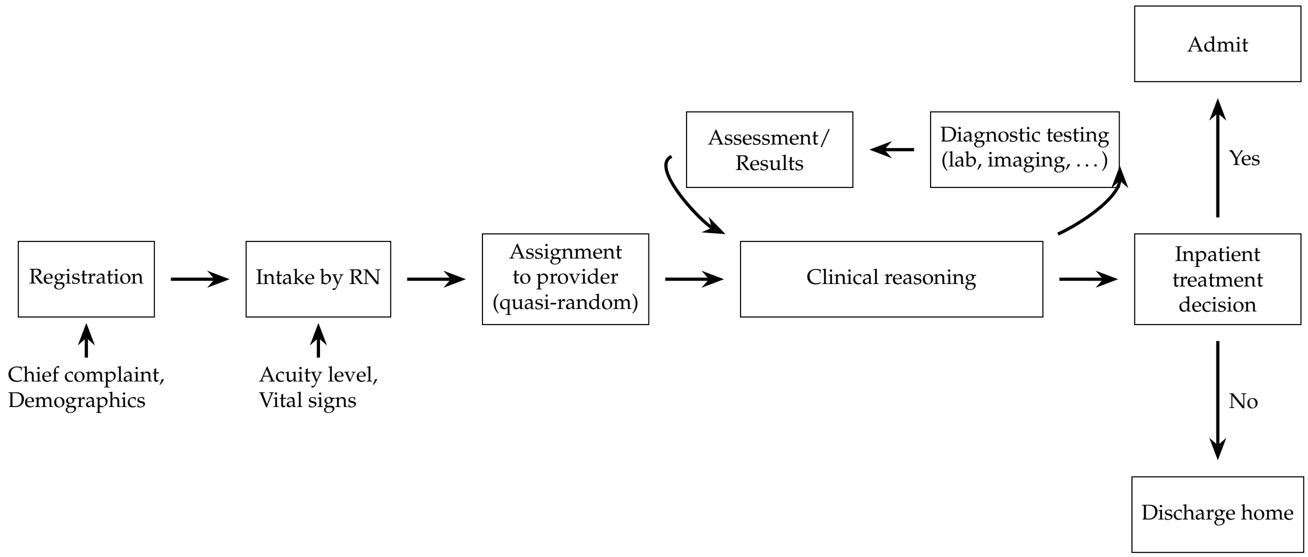


Figure 1: Emergency-department diagnostic and disposition flow.

The mean (median) patient spends slightly under 7 (5) hours in the ED. Approximately 76% of patients are discharged from the hospital at the end of their ED stay, while 24% are admitted to inpatient. The five most common patient chief complaints at intake (for the most granular tier) are abdominal pain (10.2%), chest pain (6.5%), shortness of breath (5.7%), a fall (2.9%), or altered mental status (2.8%). There is a small probability of many of the 732 possible chief complaints: we present the top 25 chief complaints, with associated probabilities, in Appendix Table D.4. Appendix Table C.2 describes the rolled up probabilities of chief complaints at the more aggregated tiers (Tier I and Tier II). 61.8% of patients are given an ESI triage score of 3, in the middle of the severity range, with 20.9% given a higher severity level of 2 of 0.4% given the highest severity score of 1. 33.1% of patients have abnormal diastolic blood pressure measured at intake, 5.7% have abnormal respiration, and 3.1% have abnormal pulse.

Table 1: Summary Statistics

	Min	Mean	Median	Max
Panel A: Encounter-level				
Demographics				
Share Male		0.505		
Share Asian		0.178		
Share Black		0.152		
Share White		0.457		
Age	18	50.863	51	89
Health status				
Share admitted to hospital within last 60 days		0.435		
Share ED visit within last 60 days		0.212		
Share upgraded to inpatient		0.238		
Share readmitted within last 30 days		0.039		
Share arrived in emergency vehicle		0.298		
Charlson comorbidity index	0.000	0.674	0.000	16.000
ED length of stay (hours)	0.017	6.981	4.933	125.100
Inpatient length of stay (days)	0.000	5.860	3.875	25.000
Chief-complaint shares				
Abdominal Pain		0.102		
Chest Pain		0.065		
Shortness Of Breath		0.057		
Fall		0.029		
Altered Mental Status		0.028		
Acuity-level shares				
Immediate		0.004		
Emergent		0.209		
Urgent		0.618		
Less Urgent		0.155		
Non-Urgent		0.012		
Missing		0.003		
Abnormal-vital-sign shares				
Abnormal systolic blood pressure		0.107		
Abnormal diastolic blood pressure		0.331		
Abnormal temperature		0.004		
Abnormal pulse oximetry		0.073		
Abnormal respiration		0.057		
Abnormal pulse		0.031		
Physician actions				

continued on next page

Table 1: Summary Statistics (*continued*)

	Min	Mean	Median	Max
Number of all orders	1	8.136	6	64
Number of diagnostic orders	0	4.552	3	47
Number of medication orders	0	2.321	2	34
Number of non-DX proc. orders	0	1.263	1	40
Number of consults	0	0.252	0	13
Time spent editing notes (in minutes)	0	22.587	13.700	379.667
N = 76,001				
Panel B: Physician-level				
Demographics				
Share Female		0.535		
Types				
Share Resident years 1/2		0.489		
Share Resident years 3+		0.270		
Share Attending physician		0.192		
Share Nurse practitioner		0.049		
Encounter-weighted types				
Share Resident years 1/2		0.419		
Share Resident years 3+		0.262		
Share Attending physician		0.145		
Share Nurse practitioner		0.174		
Shifts				
Number of shifts	1	32.505	11	390
Average shift length (hours)	5.165	11.951	10.824	29.178
Average number of patients per shift	1.000	2.919	2.278	10.585
N = 610				

Notes: Observations in Panel A are at the *encounter*-level; observations in Panel B are at the *physician*-level.

Information on the 610 medical providers is presented in bottom panel of Table 1. 53% of providers are female. 19% are attending physicians, who are more senior physicians who typically specialize in emergency medicine and who have already completed their residencies. 28.1% of providers are advanced residents, i.e., physicians in years 3 and 4 of their post-medical school training, typically also as emergency medicine specialists. 49.6% of providers are early-stage residents in years 1 and 2 of their residency, while 3.2% are nurse practitioners. The average physician in our sample has 32 shifts, with a heavily right-skewed distribution. Mean (median) shift length is 12 (11) hours and the mean (median) primary patients per shift is 2.9 (2.2), also with a right-skewed distribution. Appendix Table D.1 presents information on the number and distribution of shifts, by physician type, showing similar distributions of shifts lengths by type but each attending and nurse practitioner having more shifts on average.

The Epic Audit Log records every interaction of physicians with the Electronic Medical Record (EMR), allowing us to observe the timing (to the second) and nature of all physician actions, including placing orders (medications, lab tests, etc.), note taking, and inpatient admission decisions. The encounter-level part of Table 1 describes the typical occurrence of different EMR actions for an encounter. For example, mean number of orders per encounter is 8.1, mean diagnostic orders is 4.5, mean number of consults with other expert physicians is 0.25, and mean time spent editing notes for a patient is 22.6 minutes.

For our primary analysis dataset, we construct a panel at the *patient-provider-action* level, where each observation corresponds to an interaction between a physician and a patient. We specify a range of actions but focus the most on placing orders, taking notes, and the patient admission / discharge decision. We note that, in many cases, orders are placed in batches, and we account for this in many of our specifications, as described in detail later. For orders, we observe granular details on the type of order placed, the results of the order if diagnostic, and when results are viewed. For notes, we observe metadata on note edit time, note length, and when notes are viewed, but do not observe the full notes themselves for privacy reasons.

3.1 Assignment of Patients to Physicians

Once intake triage process is complete, patients are either taken straight to an available bed ("direct-bed") or returned to the waiting room. Patients are placed into a queue where their position is managed by the charge nurse and displayed to all physicians on a live computerized tracking board (shown in Appendix Figure B.1). The board lists, in order of arrival within each acuity tier, the chief complaint and the elapsed waiting time.

Physicians pick up cases sequentially ("next-up") from this board, typically prioritizing the patient with the most severe triage score who has been waiting the longest. At the moment of assignment physicians observe only the chief complaint and acuity code. Overall, this workflow means that i) initial assignment of acuity level (1 to 5) and chief complaint is nurse-driven and completed before any physician becomes involved and ii) physician-patient pairings are governed almost entirely by the arrival sequence of patients and general physician capacity, rather than by any patient characteristics beyond the triage score. These features are both important for identifying the effects of cognitive load in our environment, as discussed in more depth in the next section.

3.1.1 Quasi-Random Assignment

As an input into our upcoming identification discussion, we assess the quasi-random nature of patient assignment. We do this conditioning on ex ante observable variables (time-of-day, day-of-week, chief complaint, acuity code), which we also condition on in our primary empirical analysis. We define this set of conditioning variables as $\mathbf{T}_i$. We investigate the extent to which observable (ex ante and ex post) patient characteristics, defined as $\mathbf{X}_i$, are balanced across physicians conditional on $\mathbf{T}_i$.

Our main specification uses the propensity of each physician to order diagnostic tests (intensity of treatment) as the physician trait to assess assignment balance with respect to patient characteristics.[20] We define a leave-out measure of the propensity of each physician to order diagnostic tests:

$$Z_i = \frac{1}{N_{j(i)} - N_i} \sum_{i' \neq i} \mathbf{1}[i' \in I_{j(i)}] \sum_n d_{i'n} \tag{4}$$

where $I_j(i)$ is the set of patients assigned to physician j, $N_{j(i)}$ is the total number of patient-interactions physician j has, and N_i is the number of interactions patient i has with physician j. We then run the regression:

$$Z_i = \alpha + \mathbf{X_i} + \mathbf{T_i} + \varepsilon_i \tag{5}$$

with the primary goal of showing limited variation in Z_i with respect to X_i conditional on T_i. We run this regression using all patient encounters in our sample.

Figure 2 presents the results. First, we run the same regression in panel A, but with number of orders for that specific encounter as the outcome variable, to show that the variables in $\mathbf{X_i}$ capture significant variation in intensity of patient treatment. Next, we show the results from the main balance regression in panel C. Note that, for exposition, we standardize the variables in $\mathbf{X}_i$ to have mean zero and standard deviation equal to one. At the bottom of each panel we report F-statistics for the null hypothesis that all coefficients on the elements of X_i are equal to zero and the R^2 of the vector X_i. Although $\mathbf{X}_i$ is highly predictive of the total number of diagnostic tests placed for the patient, it has far less predictive power with respect to Z_i. Further, there is no obvious relationship between the coefficients in panel A and panel C. Overall, we see statistically significant but economically small / negligble imbalance in patient characteristics, show by the small F-statistic (5.9) and small R^2 (< 0.005) in (C), while these are much larger in the regression for the number of diagnostic orders in the encounter (F of 406, $R^2 = .18.$).

[20] We choose this propensity to order diagnostic tests because (i) it is a one-dimensional measure that distinguishes important aspects of physician differences (ii) we show meaningful cross-sectional differences in this measure across physicians later in the paper and (iii) diagnostic orders placed are a primary outcome variable in our main empirical analysis.

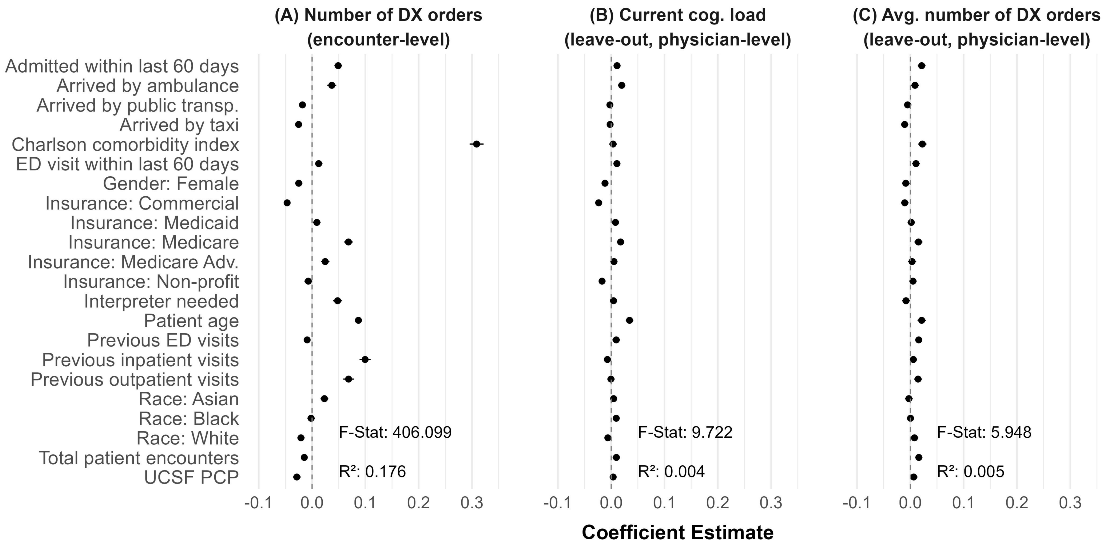


Figure 2: Covariate Balance

Notes: This figure shows coefficients and 95% confidence intervals from regressions of the total number of diagnostic orders placed for the encounter (column A), the assigned physician's leave-out current cognitive load (column B), and the assigned physician's leave-out average number of diagnostic orders placed across all encounters (column C), on covariates X_i, controlling for hour-of-day, day-of-week, hours-since-shift-start, hours-until-shift-end, acuity score, and chief complaint fixed effects. All outcome variables and covariates are standardized so that their means are equal to 0 and standard deviations are equal to 1. At the bottom of each panel, we report the F-statistic from the joint F-test of all covariates and the R^2 from the regression of the respective outcome variables on all covariates.

Panel B in Figure 2 assesses balance with respect to the point-in-time cognitive load measure used in our primary specifications, defined in the next section. We discuss this in more detail next, but note here that balance looks similar for this cognitive load measure and the Panel C leave-out order propensity measure just discussed.

We also note that, while balance tests support the use of conditional random assignment in our main regressions, most of the variation we document in cognitive load comes from the stochastic nature of patient flows into and out of the ED, rather than from assignment once in the ED. We describe this in detail momentarily.

4 Cognitive Load: Measurement and Quasi-Random Variation

A key contribution and key input into our empirical work is identifying and measuring granular point-in-time cognitive load measures for physicians. Additionally, a key goal is to establish the existence of meaningful quasi-random variation in cognitive load to support our main empirical analyses. We present our approach and address these points in this section.

4.1 Identification

Empirically, we resolve two key identification issues in order to establish the impact of cognitive load on different outcomes. These two challenges are:

1. **Endogenous Observed Outcomes and Measuring Cognitive Load.** The interim and final outcomes we observe occur for physicians with different levels of cognitive load and are therefore endogenous to that cognitive load. At the same time, we want to base our measures of cognitive load on the number of and intensity of patients a physician is treating / has treated. Consequently, we can't use actual post-physician-assignment actions as inputs into our measures of cognitive load.

   We address this issue by forming measures of predicted patient intensity using only information that is ex ante to provider assignment and treatment. We use predicted patient intensity, rather than ex post measured intensity, when forming our physician-time-specific measures of cognitive load, so that these measures, which relate to the typical mental effort used for a given patient, are not based on endogenous inputs. We describe this prediction process in detail in the next sub-section.

2. **Quasi-Random Arrival, Assignment and Variation in Cognitive Load.** The second identification challenge we address is ensuring that we have quasi-random variation in cognitive load, within and across physicians. We leverage both quasi-random patient arrival to the ED as well as quasi-random assignment of patients to physicians once they have arrived in the ED. Later in this section, we present a range of evidence showing and describing significant within-physician quasi-random variation in cognitive load over time.

4.2 Measuring Expected Workload Intensity for each Encounter

For each encounter, we construct a measure of the expected workload intensity c_i based on three outcomes: (i) the number of orders (ii) the log of note edit time (iii) the number of consultations obtained with other expert physicians. As described in the next subsection, we use these predictions of expected workload intensity as inputs into our point-in-time measures of cognitive load for each physician.

We predict each of these three variables using cross-validated LASSO regression, leveraging encounter-level variables that are ex ante to physician assignment and treatment. These predictors include patient age, patient vital signs at triage intake, the three different tiers of chief complaints (including the lowest tier with 732 possibilities), the ESI triage score, means of arrival to the ED, recent ED / admission history at UCSF, the need for an interpreter, and the Charlson co-morbidity index. Methods for dealing with sparse predictor matrices are important for our process given the importance of chief complaints in prediction. Appendix C.1 describes our LASSO procedure in full detail including myriad sensitivity checks, measures of goodness of fits, and split-sample tests to ensure we are not overfitting our predictors. Additionally, we analyze alternative prediction specifications, i.e. gradient-boosted trees and neural networks, finding that LASSO delivers similar results in our context.

Appendix Tables C.5-C.7 show the LASSO coefficients of the 25 most important predictors for each of the three outcome variables. ESI triage acuity codes and distinct chief complaints are important predictors for the all three outcome variables, with chief complaints impacting each outcome variable in sensible ways based on the complaint and outcome variable. Though, as expected, there are some predictor differences across these three distinct outcomes, predictions are highly correlated across them.

For each encounter we combine the three predictions into a composite measure c_i by summing the standardized predicted values. This composite score is further normalized to a z-score. Finally, for ease of interpretation, we apply a min-shift by adding the absolute value of the minimum composite score, producing a "patient complexity" score that is both centered and non-negative across the sample. Figure D.3 shows the distribution of patient complexity measures, which has two peaks and is right-skewed.

4.3 Cognitive Load Definition

We use these patient intensity predictions as inputs into our measures of physician point-in-time cognitive load. We compute point-in-time cognitive load for physician j taking an action for patient i at time t as follows: We define cognitive load as the sum of the encounter-level complexity measures $c_{i',j,t}$ of all distinct patients excluding the focal patient i ($i' \neq i$) for whom the physician has taken an action (i.e., orders, notes, admission decision) during the preceding 90 minutes within the same shift. We show in Appendix F.4 that all main results are robust to this assumed time window, and hold, e.g., when using instead a 60-minute or 120-minute window.[21] Note that, since we exclude the focal patient i, this leave-out measure varies across patients i even if an action would have been conducted at the same time t.

Physician j's cognitive load $F_{i,j,t}$ when attending to patient i at time t is thus given by:

$$F_{i,j,t} = \sum_{i' \in \mathcal{I}_{j,t} \setminus \{i\}} c_{i',j,t} \tag{6}$$

where $\mathcal{I}_{j,t}$ is the set of patients with whom the physician j has interacted at least once within the last $t - 90$ minutes in the current shift. Note that, if physician j has been on shift for less than 90 minutes as of time t, $\mathcal{I}_{j,t}$ comprises all patients with a physician action since the beginning of the shift.

[21] Additionally, we find that this definition is robust to including patients who the physician is assigned to, who are still in the ED, but for whom no action is taken in the prior 90 minutes.

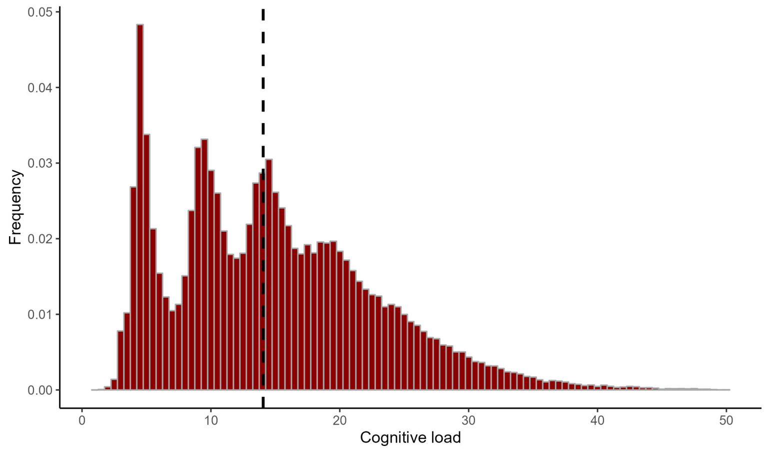


Figure 3: Distribution of Cognitive Load

The figure shows the relative frequencies of the cognitive load measure with a binwidth of 0.5. Each observation is the cognitive load at one *encounter-order action*. The vertical dashed line indicates the median.

Figure 3 presents the distribution of our measure of point-in-time cognitive load across all patient order actions in our main sample (266,625 observations). This shows significant variation in cognitive load, and also highlights the right-skewed nature of this distribution.

4.4 Variation in Cognitive Load

Figure 4 illustrates the variation of point-in-time cognitive load along several dimensions. Each figure reports different measures related to cognitive load (y-axis) as a function of hour-in-shift for a physician (x-axis).

Panel A in Figure 4 presents the distribution of within-physician averages for cognitive load as a function of shift-hour. Thus, for a given physician, the figure averages point-in-time cognitive load for a given shift hour and reports the values for each physician (with a physician-specific line) across all shift hours. For expositional purposes, the figure shows this for the top 20 advanced stage residents (by number of patients). As expected, the figure shows a tight distribution of these average values at each hour across physicians, highlighting the quasi-random nature of patient assignments within shift and over time.

We decompose variation in cognitive load within and across physician to highlight the stochastic nature of this variation. For example, during the fifth hour the shift, physicians have an average concurrent cognitive load of 8.60, with a total standard deviation of 4.76. Decomposing this overall variation via the Law of Total Variance[22], dispersion in this hour is driven largely by within-physician fluctuations – 76.2% of the total variation – while systematic differences across

[22] $Var(F) = Var(E[F \mid J]) + E[Var(F \mid J)]$, where the first term represents the *between*-component, and the second term the *within*-component.

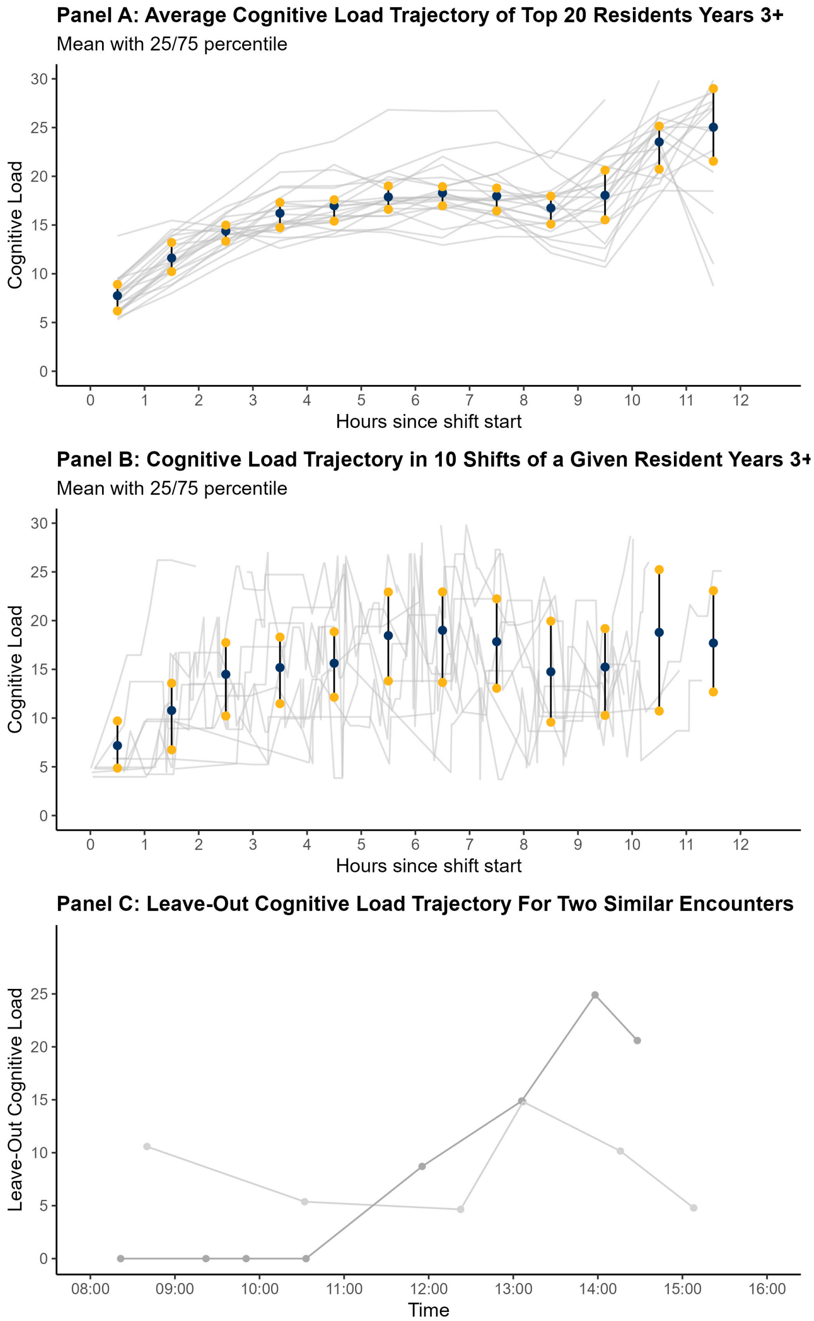


Figure 4: Trajectory of Cognitive Load

Notes: Panel A shows the average cognitive load at each shift hour for the 20 providers with the most shifts who are in year 3 or later of their residency. The blue dot denotes the average while the bottom yellow dot denotes the 25-th percentile and top yellow dot denotes the 75-th percentile of cognitive load across all shifts.

Panel B shows the cognitive load for 10 randomly selected shifts of a *given* resident in year 3 or later. These shifts are selected from the 37 shifts for that provider that last between 10 and 12.5 hours.

Panel C traces the *leave-out* cognitive–load faced by the treating physician during two abdominal-pain encounters that began on a Friday between 08:00 and 09:00 and were both triaged as *Urgent*. In each case the provider was a resident in the second hour of a 10-hour shift, yet the intensity of concurrent workload diverges sharply. The solid line shows the cognitive load for the focal physician after removing the focal patient from the calculation. Dots indicate order “batches.”

physicians account for only 23.8% of the variation. This relationship holds across shift hours. Additionally, much of the systematic variation present is due to effects like day-of-week or time-of-day, which we condition out in our primary analyses. This strongly underscores the main identifying assumptions in our regressions, namely that cognitive load is random conditional on the many control variables included.

Panel B highlights this further with an illustrative example of cognitive load for 10 shifts of one advanced resident. This panel highlights typical across-shift variation in cognitive load for a given physician, which is notably larger than any systematic differences noted in the statistics above and in Panel A.

Panel C provides an even more granular example, highlighting the leave-out cognitive load trajectories for two ex ante similar patients being treated by one resident on different shifts. The two patients come in at the same time of day and day of week, have the same chief complaint and acuity triage scores. Thus, the conditioning variables for our primary regressions are the same for these two patients, with the only differences being changes to the cognitive load of the same treating physician. The figure highlights quite different leave-out cognitive load for the physician over the course of each patient encounter, highlighting the exact variation we use in our upcoming analyses.

We also assess how variation in cognitive load comes from two proximate determinants – patient head-count versus case-mix complexity. We find that 86.3% of the spread is explained by the number of simultaneous patients, whereas 13.7% stems from how clinically demanding those patients are. This highlights that the stochastic nature of the number of patients flowing into or out of the ED is a key factor underlying variation in cognitive load in our setting, while variation in case complexity conditional on number of patients provides some additional variation.

4.5 Prior Cognitive Load

We also measure prior cognitive load from earlier in a shift in order to aggregate cognitive-load effects. Here, we sum up the encounter-level complexity measures of all distinct patients (excluding the focal patient i) for whom the physician has placed orders from the start of the shift up to 90 minutes before the current interaction:

$$G_{i,j,t} = \sum_{i' \in \mathcal{I}_{j,t-90} \setminus \{i\}} c_{i',j,t} \tag{7}$$

where now the sum is over all patients with whom physician j has interacted between the shift start and $t - 90$ minutes before the focal interaction with patient i.

We note that there are a variety of ways to account for prior cognitive load, including (i) measures that integrate current cognitive load with some depreciated notion of past cognitive load, or (ii) time-series lags for prior 90-minute windows. Additionally, we note that we could also incorporate longer-duration measures related to cognitive load experienced within other shifts over the past week or month. Finally, in a coarse way, one can think of the prior cumulative cognitive load measure as a reflection of behavior dynamics that occur when cognitive load shifts in specific directions over time.

4.6 Discussion

We define cognitive load based on the number and intensity of patients a physician sees at a given point in time. As discussed in the cumulative measure above, these point-in-time measures can be combined over time in different ways to reflect a broader notion of cognitive load that depends on the past and present patient burden.

While the number of patients and number of actions per patient are a natural starting point for mental burden, there are a number of other possibilities that could extend this notion. First, it is plausible that even if a patient has many orders and note actions, that something about the treatment experience is very routine and, consequently, does not pose a larger mental burden then a patient with a less routine case but few expected orders or note actions would. Second, one could imagine that seeing many complex patients of exactly the same kind in one day would pose a lower mental burden than seeing many patients of similar complexity but different chief complaints / conditions. Third, it may be that certain types of patients, e.g. those with dementia co-morbidities, are especially taxing to treat regardless of the primary chief complaint.

We view these kind of extensions as additive to our approach, in the sense that they would make our approach more precise and allow testing of more subtle cognitive load hypotheses. Overall, we find it unlikely that these additions would subvert our primary assumption that treating more demanding patients in terms of expected actions, and a larger number of those patients, leads to higher cognitive load. These kind of extensions could leverage hypothesis such as those mentioned above and those discussed in prior work (e.g., Enke and Graeber (2023) and Oprea (2024)) to study whether specific forms of potential cognitive load have meaningful impacts.

5 Empirical Approach

Our primary empirical analyses come from a series of regression models with the following form:

$$Y_{i,j,t} = \alpha + \beta f_1\left(F_{i,j,t}\right) + \gamma f_2\left(G_{i,j,t}\right) + \delta_t + \lambda_i + \theta_j + \varepsilon_{i,j,t} \tag{8}$$

where i indexes the patient, j indexes the physician, and t indexes time. $Y_{i,j,t}$ denotes an outcome of interest, $F_{i,j,t}$ denotes the leave-out measure of cognitive load defined in Equation (6), and $G_{i,j,t}$ represents the leave-out measure of long-run cognitive load defined in Equation (7). We consider specifications where $f_1(\cdot)$ is either a z-score transformation or a set of quintile indicators for the underlying variable. We primarily use z-score transformation for $f_2(\cdot)$ but do at times specify it as a set of quintile indicators.

Our approach focuses on leveraging the within-physician variation in cognitive load described in depth in the last section. To this end, we include physician fixed effects θ_j. We include encounter-level fixed effects for chief complaint and acuity score – captured by λ_i – for the twin purposes of (i) conditioning on the ex ante patient information known at the time of assignment and (ii) to isolate deviations in key outcomes due to cognitive load, above and beyond typical treatment for a given chief complaint and acuity level. We include several different time-related fixed effects including (i) time of day, (ii) day of week, (iii) time into physician shift, and (iv) time until the physician shift ends, captured by δ_t.

In most of our analyses we focus on β and interpret this coefficient as the causal effect of changes in recent cognitive load on our key outcomes of interest. The set of fixed effects allows us to isolate the effect of cognitive load conditional on other contextual factors that matter (time of day, day of week, typical physician practice style, type of patient condition / complaint). We can also use this specification to study how the impact of cognitive load compares to variation generated by the observable factors included in the fixed effects.

In our primary specifications, for simplicity we include only orders and notes from the assigned primary provider. In practice, 97% of notes and 60% of orders are input by the primary assigned provider.

6 Results

One of our primary focuses is to assess how cognitive load impacts physician diagnostic processes. To this end, many of our outcomes focus on different aspects of diagnostic orders, i.e. orders placed with a specific goal of moving towards a final diagnosis. These outcomes directly relate to four hypotheses presented in our model in Section 2.

We begin by looking at the amount of diagnostic orders placed. One comparative static from our model is that the number of diagnostic orders placed will increase as cognitive load increases, as a substitute for physician thinking effort. We use an order batch for patient i and physician j as the action (occurring at time t), and the outcome variable in Equation (8) is the number of diagnostic orders placed in the order batch.

Table 2 presents the results in columns (1) (without physician fixed effects) and (2) (with physician fixed effects). For our preferred specification with physician fixed effects, we find that a one standard deviation increase in cognitive load leads to 0.05 additional diagnostic orders in a batch, equivalent to a 4 percent increase relative to the average number of diagnostic orders. These estimates have a high degree of precision, indicative of the large number of observations in our data.

Columns (3) and (4) report the effect of cognitive load on medication orders, providing a useful

Table 2: OLS Estimates of the Impact of Cognitive Load on Orders

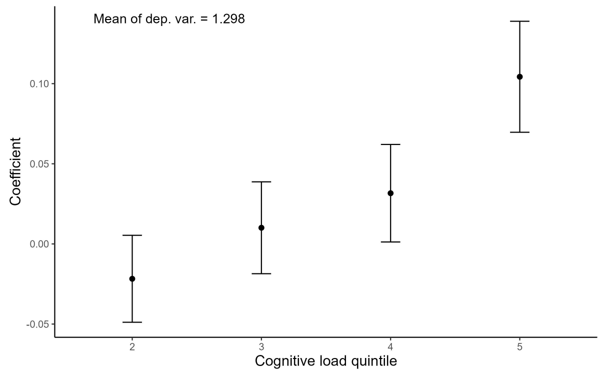
	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.071***	0.050***	0.010***	0.002
	(0.006)	(0.006)	(0.003)	(0.003)
Prior cognitive load - Control	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,625	266,625	266,625	266,625
R^2	0.043	0.053	0.024	0.057
Adjusted R^2	0.040	0.048	0.021	0.052

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

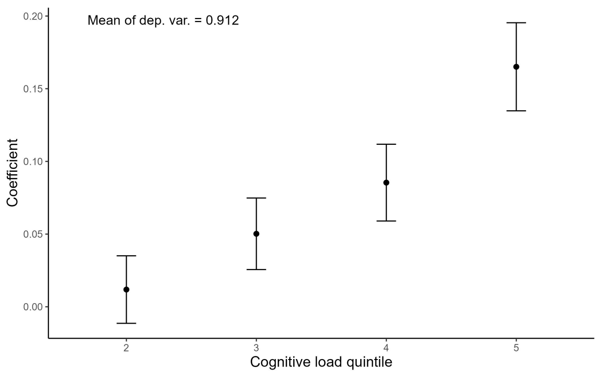
validation check since medication orders should be less subject to cognitive load and not closely related to diagnostic effort. We find that a one standard deviation increase in cognitive load has a precise null effect on the number of medication orders. This highlights how physicians are not uniformly increasing all order types and, instead, selectively expanding the number of diagnostic orders.

We note that this specification, and most that follow, occurs at the diagnostic order action level. This means that it assesses the intensive margin of orders placed in an order batch. We focus on the intensive margin because it allows us to precisely control for cognitive load at the time of order placement rather than relying on an average (or other moments) over the course of a given encounter. As described later in our robustness section, we also assess the effect of cognitive load on the extensive margin of diagnostic orders, i.e. how many distinct order batches a physician places. We find that a physician with a one standard deviation higher leave-out cognitive load, averaged over a patients full encounter, leads to 0.117 more order batches per encounter, a 3.8% increase over baseline.

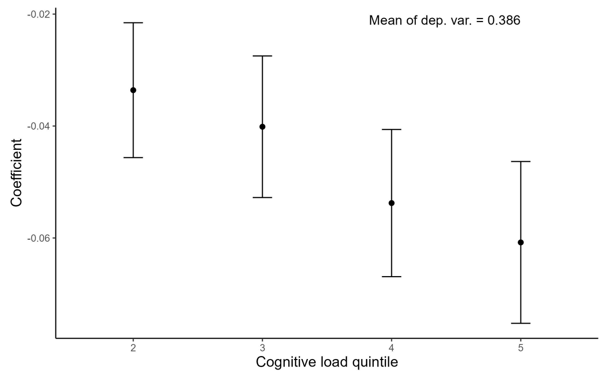
Additionally, our main specification above assumes a linear effect of cognitive load on the out-



Panel A: Number of Diagnostic Orders



Panel B: Number of top-25 Diagnostic Orders



Panel C: Number of non-top-25 diagnostic orders

Figure 5: Cogntive Load Quintile Estimates

Notes: The figure plots coefficients from linear regressions of number of diagnostic orders on quintile indicators of the leave-out measure of cognitive load. Panel A uses the number of diagnostic orders as the outcome variable, Panel B uses the number of diagnostic orders that are among the 25 most frequently placed diagnostic orders, and Panel C uses the number of diagnostic orders that are outside the 25 most frequently placed diagnostic orders. The unit of observation is the encounter-action. All specifications control for prior cognitive load and include fixed effects for the hour since shift start, the hour until shift end, hour of day, day of week, chief complaint, acuity code, and physician. The first quintile is the omitted category. Heteroskedasticity-robust standard errors are used.

come variables. Appendix Table F.2 examines a specification where $f_1\left(F_{i,j,t}\right)$ maps our cognitive load measure $F_{i,j,t}$ to a set of quintiles. Panel A in Figure 5 presents the quintile estimates for β for the total number of diagnostic orders, relative to the bottom quintile (least cognitive load).

We find an ordered, but non-linear effect, with much of the overall effect driven by physicians in the top 20% of cognitive load. When a physician is in this most cognitively loaded quintile, the number of diagnostic orders by 9% relative to the leave out quintile. As the appendix table shows, there is no effect of these higher cognitive load quintiles on medication orders.

A next natural question is to ask how the types of diagnostic orders placed change as cognitive load increases. Table 3 shows a clear shift in the profile of orders. When physicians have higher cognitive load, they place a meaningfully higher number of diagnostic orders in the set of top 25 most common diagnostic orders. In the specification with physician fixed effects (column (2)) physicians place 8% more common orders per order batch for a one standard deviation increase in cognitive load. Column (4) shows that the flipside of this is also true. A one standard deviation increase in cognitive load *decreases* the number of uncommon diagnostic orders by 5%. Later, we replicate this finding for specific common chief complaints. These findings highlight how, when physicians are cognitively loaded, they shift to a more "standardized" or "routine" set of diagnostics and move away from more targeted, less common orders, commensurate with a model where physician do less thinking / evaluation of a given patient up front.

Table F.3 in the appendix studies how the profile of orders changes with cognitive load quintiles (instead of linear in standard deviation). Panels B and C in Figure 5 present a visual representation for β for (i) common diagnostic orders and (ii) rarer diagnostic orders, respectively.

The pattern is striking: for the most cognitively loaded quintile, common (top 25) diagnostic orders increase by 18%, relative to the least cognitive loaded quintile, while uncommon orders decrease by 17%. For each outcome, the relationships are monotonic across the five quintiles as cognitive load increases, both for common diagnostic orders and uncommon ones. Additionally, in appendix section F.7, we study the threshold for common / uncommon orders, using the 10 most common, 50 most common, and 100 most common to separate these categories. We find similar patterns in these analyses to those described here.

To shed light on *which* orders account for these shifts toward more common diagnostics, Appendix Figures F.4–F.8 plot coefficients from separate linear probability models for the most frequent diagnostic and medication orders and order subcategories, focusing on two canonical chief complaints: chest pain and abdominal pain. We regress an indicator for placing a given diagnostic or medication order (or at least one order in a diagnostic or medication subcategory) on the leave-out cognitive load measure and include the same battery of fixed effects as in Table 2.

Several patterns emerge. First, higher cognitive load systematically raises the use of standard diagnostic work-ups tailored to the presenting complaint: for chest pain, the largest increases are in troponin testing, chest imaging, and related lab work. For abdominal pain, the increases appear in basic metabolic and liver panels, lipase, and abdominal CT or ultrasound. These order-specific results mirror the evidence from Table 3 that cognitively loaded physicians lean more heavily on common diagnostic tests rather than rare ones. Second, cognitive load also increases symptomatic treatment and supportive care: across both chief complaints, physicians are more likely to order analgesics, antiemetics, and intravenous fluids when load is high. Taken together, these patterns suggest that under higher cognitive load physicians respond by intensifying standardized diagnostic bundles and symptom relief.

Before we move on to our analysis of entropy and diagnostic precision, we present one more view into how medical practice changes with cognitive load. We examine documentation effort by examining the time a provider spends editing patient notes. This measure, developed from audit log data, provides time-stamped information on physician note-taking, allowing us to assess

Table 3: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.090***	0.068***	−0.019***	−0.018***	−0.174***	−0.071***
	(0.005)	(0.005)	(0.002)	(0.002)	(0.005)	(0.004)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.019	0.107

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). For columns (5)-(6), the outcome variable is log-transformed, while its mean is reported in raw seconds. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

whether physicians reduce note-taking when cognitively loaded.

Columns (5) and (6) from Table 3 reports the results from specification (8), using 181,154 note edits as the unit of observation and the natural logarithm of note editing time as the outcome. Column (6) shows that a one standard deviation increase in cognitive load reduces editing time by about 7.1 percent. This suggests that, physicians have less mental bandwidth to edit notes, though we don't assess whether or not this has positive or negative welfare effects.[23] [24]

6.1 Diagnostic Precision

We now investigate the precision of diagnostic orders, in a structural sense, as a function of cognitive load. We construct a measure of the "informativeness" of a physician's diagnostic orders

[23]As for resource utilization, our analysis focuses on the *intensive margin* of note edit time. On the *extensive margin*, a physician with a one-standard-deviation higher cognitive load has 0.05 more note-edit instances per encounter (+2.3% relative to the 2.17 baseline), partially offsetting the diminished effort on the intensive margin and highlighting a more piecemeal production of notes when cognitively loaded.

[24]Recent work by Kloiber (2025) shows that discharge notes that are more complex in terms of language lead to higher post-discharge mortality. This is suggestive of the types of downstream impacts from note modifications, though we do not study note language complexity or discharge notes

with respect to possible patient diagnoses, thereby recovering the physician's implied prior over diagnoses from the diagnostic orders placed. This structural measure builds on prior work in information theory (e.g., Caplin and Dean (2015)), and links directly to a comparative static in our model stipulating that, when physicians are more cognitively loaded, they exert less mental effort on a given patient up front and start placing diagnostic orders with more diffuse priors.

To construct our informativeness measure, we proceed in two steps. In the first step, for each diagnostic order o_z and chief complaint cc, we estimate the empirical probability mass function of final diagnoses among encounters in which order o_z is placed for patients presenting with chief complaint cc:

$$\widehat{\Delta}_{DX}\left(\cdot \mid o_z, cc\right) \quad \text{with elements} \quad \widehat{\Delta}_{DX}\left(d \mid o_z, cc\right) \ \text{ for each final diagnosis } d \in \mathcal{D} \tag{9}$$

Next, let $O_i = \{o_{i,1}, \ldots, o_{i,z}, \ldots, o_{i,Z_i}\}$ denote the set of diagnostic orders placed for encounter i. To rule out dynamic considerations such as physician-learning from observing test results, we focus on orders placed within the first 30 minutes of the initial patient–physician interaction. Since we are underpowered to estimate the *joint* implied distribution over diagnoses for an order set O_i, we construct an "aggregate" implied distribution by summing and then re-normalizing the marginal distributions of each order in the order set:

$$\widehat{\Delta}_{DX}\left(\cdot \mid O_i, cc_i\right) = \frac{1}{Z_i}\sum_{z=1}^{Z_i}\widehat{\Delta}_{DX}\left(\cdot \mid o_{i,z}, cc_i\right) \tag{10}$$

We then compute the entropy of this aggregate distribution:

$$H_i = H\Big(\widehat{\Delta}_{DX}\left(\cdot \mid O_i, cc_i\right)\Big) = -\sum_{d\in\mathcal{D}}\widehat{\Delta}_{DX}\left(d \mid O_i, cc_i\right)\log_2\Big[\widehat{\Delta}_{DX}\left(d \mid O_i, cc_i\right)\Big] \tag{11}$$

Here, $\widehat{\Delta}_{DX}\left(d \mid O_i, cc_i\right)$ denotes the conditional empirical probability of diagnosis d given chief complaint cc_i and order set O_i, as implied by $\widehat{\Delta}_{DX}\left(\cdot \mid O_i, cc_i\right) = \frac{1}{Z_i}\sum_{z=1}^{Z_i}\widehat{\Delta}_{DX}\left(\cdot \mid o_{i,z}, cc_i\right)$. Under this measure, a higher entropy value H_i implies an order set that is associated with a more dispersed prior over possible final diagnoses.

Next, we return to the ex ante phase and compute the entropy of the empirical distribution of final diagnoses associated with each chief complaint cc_i. This provides a baseline entropy for each chief complaint that is reduced as thinking occurs and orders are placed. Define this ex ante chief-complaint entropy as $H\Big(\widehat{\Delta}_{DX}\left(\cdot \mid cc_i\right)\Big)$. We define the entropy reduction from physician assessment and contemplation, prior to placing the first order batch, as:

$$E_i := H\Big(\widehat{\Delta}_{DX}\left(\cdot \mid cc_i\right)\Big) - H_i \tag{12}$$

Here, E_i measures the baseline entropy associated with chief complaint cc_i minus the updated

Table 4: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision

	Entropy reduction (cond. on CC; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.024^{***}	-0.014^{***}	-0.011^{**}	-0.009^{**}
	(0.004)	(0.005)	(0.005)	(0.004)
Number of DX orders			-0.049^{***}	-0.037^{***}
			(0.001)	(0.001)
Share of top100 DX orders				-1.767^{***}
				(0.031)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.62	0.62	0.62	0.62
N	56,398	56,398	56,398	56,398
R^2	0.492	0.506	0.534	0.640
Adjusted R^2	0.485	0.494	0.523	0.631

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the encounter, considering all orders placed within 30 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Heteroskedasticity-robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

entropy H_i, which reflects physician assessment and contemplation prior to placing the first order batch. We use these measures as outcomes in our main specification to assess the impact of cognitive load on mental effort and diagnostic precision.

Table 4 shows the results for the impact of cognitive load on entropy reduction E_i. This analysis focuses on the 56,398 encounters for which any diagnostic orders are placed within 30 minutes of the first physician-patient interaction.[25] The results in column (2) imply that a one standard deviation increase in cognitive load leads to a 2.3% decrease in the entropy reduction, holding all else equal. Put differently, physicians with higher cognitive load are less precise in placing relevant diagnostic orders, i.e., placing less effort in contemplation pre-order-placing. A 2.3% decrease in entropy reduction at the mean given in Table 4 (0.62 bits) corresponds to about 0.014 fewer bits of

[25] In the appendix, we also assess the impacts of cognitive load on entropy reduction with different initial time windows and different sets of conditioning/ex-ante factors, finding similar results.

information, making the posterior distribution roughly $2^{0.014} \approx 1.010$ times more diffuse, which is about a 1% increase in the effective number of plausible diagnoses.

Column (3), our preferred specification, adds a control for the number of diagnostic orders placed, to account for the mechanical effect that a larger order set is informative over more diagnoses. Since our earlier results show that cognitive load increases the number of orders placed, we believe this is a valuable control. The coefficient of interest shows only a slightly smaller effect than that in column (2), suggesting that the reduction in diagnostic precision among cognitively loaded physicians is not driven by these physicians simply placing more diagnostic orders.[26] Overall, these results highlight how cognitively loaded physicians are less precise in their diagnostic orders, holding many factors constant (including the physician and chief complaint).

Table F.4 presents the estimates from the specification with quintiles of cognitive load, being in the top quintile of cognitive load leads to a 5% decrease in entropy reduction relative to the bottom quintile of cognitive load. As with other outcomes, much of this change occurs moving from the fourth to the fifth (top) quintile of cognitive load, illustrating the non-linear relationship between our cognitive load measure and the targeting of diagnostic tests.

6.2 Hospital Admissions

One key outcome from a visit to the emergency department is whether the patient is admitted to the hospital as an inpatient. This is a very consequential and costly decision, often leading to a meaningfully longer and more intensive patient visit. Our model in 2 generates the hypothesis that, if physicians are cognitively loaded, they will be more likely to admit patients to inpatient.

We investigate this empirically by modifying our main regression specification such that an observation is at the encounter level rather than at the patient-event level. Since the admission decision is a product of many things occurring during the encounter, it is appropriate to treat this outcome as reflecting the many things that happen during the ED encounter. The primary change moving to the encounter-level specification is that, instead of using a point-in-time cognitive load measure, we assess cognitive load over the entire patient encounter and use the maximum of cognitive load during that time as our variable of interest.

Table 5 presents the results. Strikingly, we find that a one standard deviation increase in maximum cognitive load for the physician during the patient encounter increases the chance of admission by 9%, with a very high level of statistical significance. As with earlier estimates, this controls granularly for physician, chief complaint, acuity code, time, and shift hour fixed effects. Figure 6 presents the results of our specification that splits cognitive load into deciles and assesses the impact of those deciles on hospital admissions (coefficient estimates are presented in the Appendix

[26] Column (4) adds the share of diagnostic orders in the 100 most common diagnostic orders to account for the potential mechanical effect that more frequently placed orders are associated with a wider array of diagnoses. The coefficient of interest decreases minimally, and we don't prefer this specification because controlling for more or less common orders in any way has a clear relation to our measure of diagnostic precision.

Table 5: OLS Estimates of the Impact of Cognitive Load on Inpatient Admission and Readmission at the Encounter-Level

	Inpatient admission		Readmission within 30 days	
	(1)	(2)	(3)	(4)
Max cognitive load (std.)	0.023***	0.021***	0.001	0.001
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.24	0.24	0.04	0.04
N	76,001	76,001	76,001	76,001
R^2	0.221	0.242	0.048	0.062
Adjusted R^2	0.213	0.228	0.039	0.045

Notes: The table presents OLS estimates of linear regressions of dummy variables whether the patient was admitted to the hospital during the encounter or re-admitted within 30 days on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

in Table G.4). The likelihood of inpatient admission increases significantly as cognitive load increases over the deciles. The relationship is near monotonic, and shows that, all else equal, when a physician is in the highest decile of cognitive load they are 28% more likely to admit the patient than when they are in the lowest decile of cognitive load. We also assess the impact of cognitive load on readmission to the ED within 30 days (columns (3) and (4)). We find a precise null effect on readmissions.

The admissions result is notable and bears further discussion. One possible reason for the admission increase is that the ED bed capacity is met when many patients are in the ED and and there is now an additional shadow value of admissions coming from that capacity constraint. The UCSF ED operates with flexible capacity rather than a fixed bed count: during our study period, it comprises 35 patient rooms plus about 25 hallway bays, with observation and triage areas opened or closed as needed. We assess the number of patients in (and moving through) the ED during periods when physicians have high cognitive load. To do this, we plot the total number of patients in the ED when physicians have an order action for a patient and are categorized as being in the top two quintiles of cognitive load. Appendix Figure D.4 plots this distribution. The figure highlights multiple reassuring findings. First, the total number of patients in the ED at one

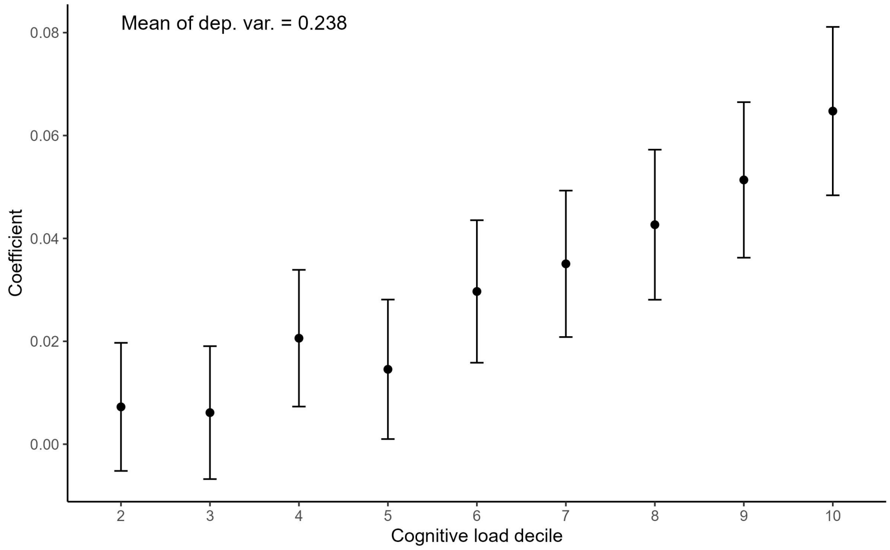


Figure 6: OLS Estimates of the Impact of Decile of Cognitive Load on Inpatient Admissions
Notes: The figure plots coefficients from linear regressions of a dummy variable whether the patient was admitted to the hospital during the encounter on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter. The unit of observation is the encounter. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, acuity code, and physician. The first decile is the omitted category. Estimates correspond to Column (4) of Appendix Table G.4. Robust standard errors are used.

time (including those not assigned to a bed yet) is typically well below capacity with only 9.66% of observations occurring when the ED is over 80% capacity. Second, and perhaps more importantly, the distribution of patients in the ED for physician actions in the top two quintiles of cognitive load looks very similar to the distribution for the lower three quintiles, suggesting that these kind of structural capacity issues do not cause these results.

A potential model-based explanation for our admissions finding is physician risk-aversion that manifests in cases with high cognitive load. Our entropy reduction results above highlight how physicians with higher cognitive load do less contemplation up front and place less targeted orders for patients, all else equal. In Table G.3 in the Appendix, we highlight how this result also holds at the encounter level: i.e., when a physician has higher max cognitive load over a patient encounter, entropy reduction over the whole encounter is lower.[27] Thus, one potential underlying mechanism is that physicians have more coarse information about a patient when they are cognitively loaded and, as a result, they are more likely to admit the patient because they are risk averse about the (higher) probability of a negative outcome if they discharge the patient instead. We also discuss this issue in the context of our model in Section 2.

[27]Note that as in our primary results, though entropy reduction is lower when max cognitive load is higher, the absolute number of orders is also higher in this case (Table G.1). This continues to suggest that, under higher cognitive load, physicians place more, less targeted, orders.

6.3 Characterizing Marginally Admitted Patients

Higher cognitive load causes increased hospital admissions. A natural next question is: are the marginally admitted patients due to higher cognitive load more or less severe cases than those typically admitted? If physicians indeed have higher case uncertainty when they are cognitively loaded and are risk averse, they may admit less severe patients on the margins to be sure that they receive sufficient care. We examine this hypothesis from an ex ante point of view primarily (using intake measures) but also assess the implications for some ex post outcomes.

We run an analysis in the spirit of Abadie (2003) to characterize the marginal admissions (in this case, patients admitted due to having a treating physician with higher cognitive load). We use cognitive load as an instrument for inpatient admission, and then back out the traits of the compliers. Let X_i denote the patient characteristics, T_i an indicator of inpatient admission, and Z_i the cognitive load of the provider treating patient i.

We estimate the IV system, using Z_i to instrument for T_i:

$$\begin{aligned} &\text{(2SLS)} \qquad X_i \times T_i = \alpha + \gamma \times T_i + u_i \\ &\text{(FS)} \qquad T_i = \lambda + \pi \times Z_i + v_i \end{aligned}$$

Abadie (2003) shows that, in the canonical binary-instrument setting, the IV estimate γ recovers the average of X_i among compliers, i.e. the average among those induced into treatment (inpatient admission) by the instrument. A closely related characterization result for general design-based IV specifications, including settings with non-binary treatments or instruments, is given in Hull (2025). In our setting, this provides a useful perspective for interpreting the resulting estimates as characteristics of the effective population induced into admission by higher cognitive load.

We implement this approach for a range of patient characteristics. We use the maximum of the standardized leave-out cognitive load measure over the course of the patient encounter as the instrument. In all specifications, we include fixed effects for the maximum shift hour during which the physician interacted with the encounter, physician, chief complaint, and acuity level, (with the relevant fixed effect omitted when the characteristic being characterized is itself an acuity measure), as well as a control for the maximum cumulative long-run cognitive load (not shown in the regression equations for expositional simplicity). We also run specifications with different forms for the cognitive load measure and find similar results.

Figure 7 presents the results. For the outcomes listed on the y-axis, we present the mean estimates for three groups (i) all admitted patients, (ii) marginally admitted patients due to high cognitive load and (iii) all patients discharged from the ED without being admitted to inpatient.

The results paint a clear picture: the patients who are marginally admitted due to higher physician cognitive load are healthier on average from an ex ante point of view and have better out-

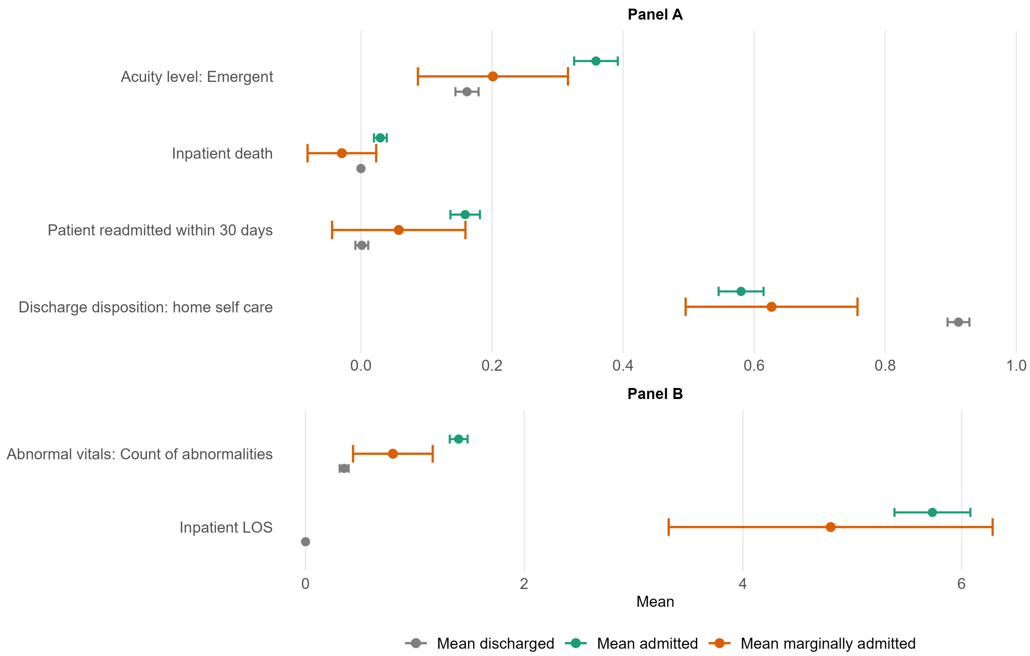


Figure 7: Characterization of marginally admitted

Notes: The figure plots the control-adjusted means for discharged encounters (gray) and admitted encounters (green), alongside an estimate of the mean characteristic for the *marginally admitted* (orange) with 95% confidence intervals. We implement a complier-characterization approach in the spirit of Abadie (2003) using a continuous instrument: the maximum of the standardized (leave-out) measure of cognitive load. Marginally-admitted means are obtained from 2SLS specifications with fixed effects for the maximum shift hour during which the physician interacted with the encounter, physician, chief complaint, and acuity level (with the relevant fixed effect dropped when the plotted characteristic is itself an acuity code), and a control for the maximum cumulative long-run cognitive load. Heteroskedasticity-robust standard errors are used.

comes ex post as well. For ex ante characteristics, marginal admits are about half as likely to have the more severe triage acuity score (Emergent) relative to the average admitted patient, and have many fewer abnormal vital signs at ED intake (about 50% lower than mead admitted). In terms of ex post outcomes (which can be impacted by differential care received in the ED), marginally admitted patients are much less likely to die during their inpatient stays and are also much less likely to be readmitted within 30 days. All of the above reflect statistically different mean characteristics, at a 95% level. Marginally admitted patients also have 20% shorter inpatient lengths of stay than the mean admitted patient on average, though this is not statistically different due to high SEs on the mean for marginally admitted patients.

It is important to note that this characterization assumes that cognitive load is a valid instrument for inpatient admission. The more that ED care differs due to cognitive load the more that patient health may differ, conditional on Xs, at the time of inpatient admission. With that said, the fact that *both* the ex ante and ex post characteristics show that marginal admits are meaningfully less severe cases highlights the validity of the instrument and the results that marginally admitted patients are healthier than the typically admitted patients.

6.4 Additional Results

As we presented our results, we highlighted a number of additional specifications and robustness checks in the appendix. These include (i) alternative functional forms for cognitive load in our main regression specification (e.g., indicators for quintiles of cognitive load), (ii) different functional forms for outcome variables (e.g., logs vs. levels), (iii) alternatives to our 90-minute window for defining action-level cognitive load (60 minutes and 120 minutes), (Iv) specifications that change the threshold for defining common vs. uncommon orders and (v) a range of specifications that are based at the encounter level, rather than the encounter-action level. We highlighted these specifications when relevant, in comparison to our results earlier, finding that they generally support our primary conclusions and expanding our understanding of the effects in the ways described.

6.4.1 Specific Chief Complaints

We also study how our analysis applies to specific common chief complaints, to get a more granular assessment of our main findings. The detailed results are presented in Appendix F.8. For this analysis, we study two common chief complaints where diagnosis is crucial, abdominal pain and chest pain. For abdominal pain, a chief complaint for 9% of the encounters in our main sample, we find very similar results to our primary findings. A one standard deviation increase in cognitive load increases diagnostic orders by 4%, does not increase the number of medication orders, meaningfully increases the number of common orders (9%), meaningfully decreases the number of rare orders (12%), decreases note edit time (9%), and reduces entropy reduction in the first order batch by 3% (though this latter effect is marginally not statistically significant). The impact on admissions is large for the top two deciles of cognitive load: there is a 50% increase in admissions moving from the top decile to the bottom decile of cognitive load, and a 35% increase moving from the second top decile to the bottom decile. All other deciles are not statistically different from the bottom decile and have smaller point estimates.

For chest pain, a chief complaint for 4% of the encounters in our main sample, we also find very similar results to our primary findings. A one standard deviation increase in cognitive load increases the number of diagnostic orders by 8%, slightly *decreases* the number of medication orders (3%), meaningfully increases the number of common orders (10%), decreases the number of rare orders (4%, marginally not statistically significant), decreases note edit time (8%), and reduces entropy reduction in the first order batch by 3% (though this latter effect is marginally not statistically significant), and does not have enough statistical power to detect an entropy effect below a 10% change. There are no clear statistically significant patterns for admission of chest pain patients with respect to cognitive load, though the standard errors are big due to lower sample size.

These results for individual chief complaints are consistent with our findings aggregating over all chief complaints. They open a window into future work that leverages structural differences in the medical practice of treating these distinct chief complaints to (i) model diagnostic pathways in a more sophisticated way, (ii) assess the impacts of cognitive load on those pathways, (iii) assess heterogeneity in physician practice styles in a granular way and (iv) assess the interactions between those practice styles and cognitive load. To answer some of these questions effectively, researchers will need an even larger sample of patients than that which we have here, in order to gain statistical power and link structural insights across different categories of chief complaints.

6.4.2 Health Outcomes

Additionally, we investigate specific higher-level health-related outcomes including the impact of cognitive load on (i) ED death, (ii) inpatient death, (iii) inpatient length of stay and (iv) 30-day readmission. The results for the specification with cognitive load deciles are presented in Appendix G.3. We find that there is limited statistical power to detect effects for any of these outcomes and cannot form any clear conclusions on these relationships.

6.4.3 Provider Heterogeneity: Cognitive Load Impacts

A natural next question for our analysis is whether certain types of physicians are more impacted by cognitive load than others. Physicians with certain practice styles, physicians with different levels of experience, and physicians with different social characteristics could be impacted differentially by cognitive load.

Our sample has limited statistical power to robustly detect heterogeneous cognitive load effects. In order to assess what heterogeneous effects could be present, we use causal forests to estimate heterogeneous treatment effects, following the work of Wager and Athey (2018) and Athey et al. (2019). This approach provides a thorough and unstructured way to assess whether there are heterogeneous treatment effects that we have enough statistical power to detect.

Our analysis incorporates a range of provider characteristics X_j including, e.g., provider category at UCSF, specialty of provider, gender, time at UCSF, medical school ranking, and date of medical school graduation. We use these characteristics as moderators of the effect of cognitive load on our key outcomes from earlier in the paper, including the number of diagnostic orders, the composition of orders (common vs. rare), entropy reduction, and note edit time.

Appendix H details our analysis. We estimate a specification of the form:

$$\tilde{Y}_{ijt} = \tau(X_j)\tilde{W}_{jt} + \varepsilon_{ijt},$$

Y_i denotes our outcome, W_i denotes the measure of contemporaneous cognitive load and $\tau(\cdot)$ is an unknown function capturing heterogeneity in the marginal effect of cognitive load intensity. For each outcome and sample, we fit a causal forest using the residualized variables $(\tilde{Y}_i, \tilde{W}_i)$ and moderators X_i.

In the Appendix, we report results of the average treatment effects of cognitive load from this specification, and detailed heterogeneity estimates using the best linear projection (BLP) of the causal forest estimates. The average treatment effects replicate our primary results, as expected. While some of the heterogeneous treatment effects are statistically different from zero, the standard errors of the effects are large in general and there is no meaningful evidence on net for a particular form of heterogeneous effects. These results suggest that our data don't have sufficient statistical power to study heterogeneous effects of cognitive load in a compelling way. As we discuss in the conclusion, we believe that performing this analysis with a meaningfully larger administrative data sample with similar depth, or an experimental approach focused on heterogeneity, are fruitful avenues to assess heterogeneous effects.

6.5 Patient Reallocation Analysis

One possible implication of our findings is that, if patients could be allocated to providers with lower cognitive load, we could reduce hospital admissions and reduce the number of diagnostic orders, while actually increasing the informativeness of those orders. In this section, we analyze this possibility, while being cognizant of the fact that we hold many contextual and dynamic factors fixed in order to do so. In that sense, we view this counterfactual analysis as suggestive of the potential for patient reassignment in a world with cognitive load, rather than a full treatment of this problem.

To do this, we leverage estimates from regression specifications at the encounter level, similar to those used earlier in our section on inpatient admissions. As discussed earlier, in this specification cognitive load for an encounter is assumed equivalent to the maximum cognitive load that occurs at a point-in-time during the encounter. Relative to our main event-action point-in-time regressions, this allows for an extension of our framework to the encounter level, while somewhat reducing the closeness in time of the cognitive load measure to the regression outcome of interest.

Figure G.1 presents the impacts of max cognitive load deciles on the number of diagnostic orders placed (regression coefficients presented in Table G.4). This is an analogous figure to Figure 6, presented earlier for the inpatient admission outcome. The number of diagnostic orders placed is monotonically increasing in the cognitive decile, with a physician in the top decile placing 42% more diagnostic orders for the patient than when they are in the lowest decile of cognitive load, all else equal. Using these results as a basis, we simulate a scheduling algorithm in which arriving patients are reallocated to the physician exhibiting the lowest current cognitive load. In the event of ties, patient assignment is determined randomly. This counterfactual policy is motivated by our earlier results demonstrating that increased cognitive load is associated with a convex increase in order placement costs. By redistributing patients so as to balance these cognitive demands more evenly across physicians, the algorithm is designed to reduce the aggregate cognitive burden

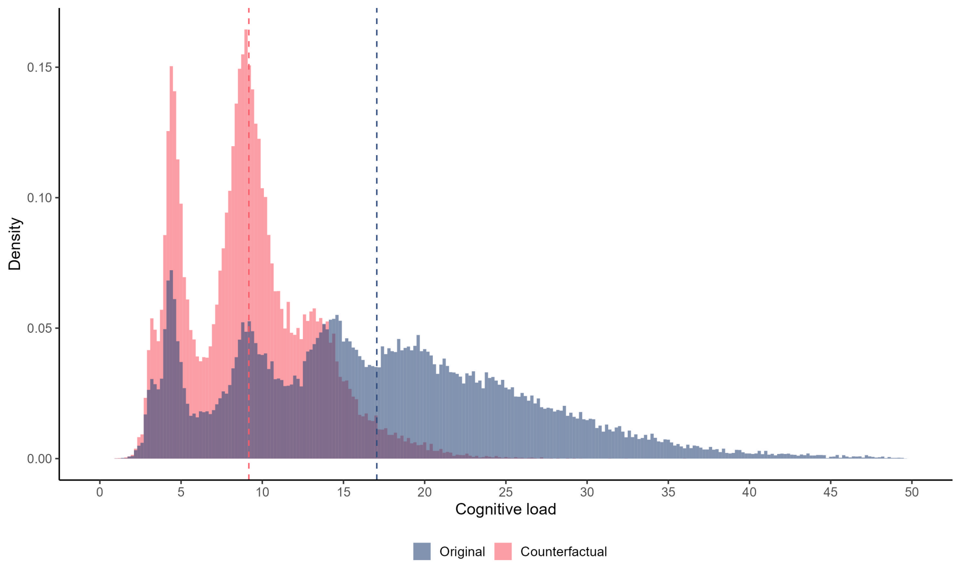


Figure 8: Distribution of Cognitive Load Under Actual and Counterfactual Patient Allocation

Notes: This figure shows histograms of the actual cognitive load (blue) and the cognitive load under the counterfactual allocation where patients are assigned to the physician on shift with the lowest cognitive load (red). The dashed lines indicate the means under the respective regimes.

during high-intensity periods.

Figure 8 displays the kernel density estimate of cognitive load under actual versus counterfactual patient allocations. The counterfactual distribution is shifted substantially toward lower cognitive load values, indicating that the reallocation effectively alleviates peak cognitive pressure on physicians. Interestingly, the counterfactual distribution has two peaks, likely corresponding to times when the ED is busy vs. times when the ED is not busy. Still, it is notable that, even using this fairly coarse mechanism, patients under this counterfactual assignment are treated by providers with much lower cognitive load.

Appendix Table G.10 presents the transition matrix of physician assignments resulting from reassignment based on cognitive load. One might be concerned that the algorithm significantly shifts assignments from one type of provider to another, conflating reductions in cognitive load with changes in provider type. Reassuringly, the table shows a stable transition matrix, with similar assignment shares within provider category in actual vs. counterfactual assignments. The off-diagonal values are also not large, suggesting no systematic cross-substitution at a granular level.

Figure 9 illustrates the differences in the distribution of diagnostic orders under the counterfactual reassignmet relative to actual observations. This figure highlights how, as a result of the compression in the cognitive load, the order distribution also becomes more compressed and, on average, lower in terms of the mean.

Table 6 presents the average impacts of the reassignment on the key outcomes of (i) number of

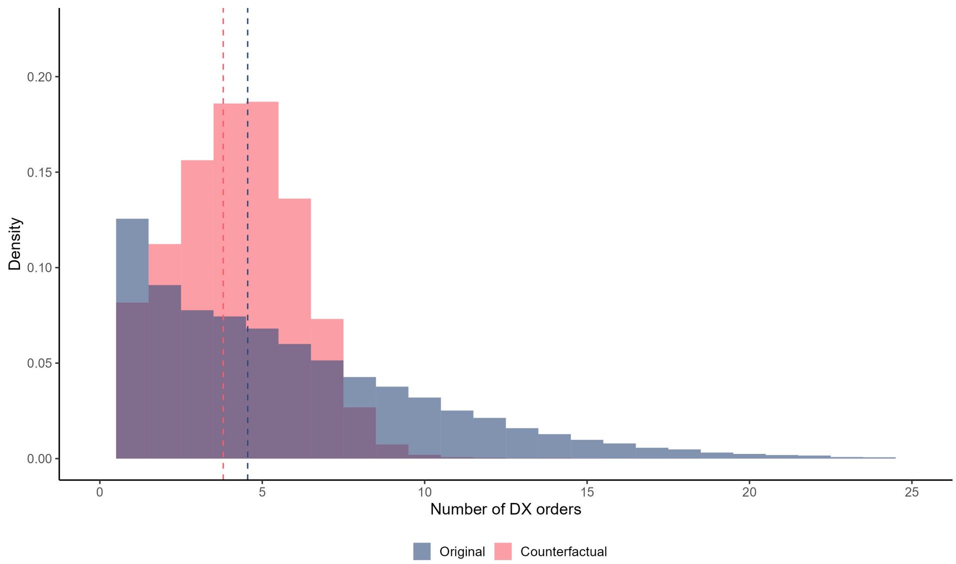


Figure 9: Distribution of Number of Diagnostic Orders Under Actual and Counterfactual Patient Allocation

Notes: This figure shows histograms of the actual number of diagnostic orders placed (blue) and the number of diagnostic orders under the counterfactual allocation where patients are assigned to the physician on shift with the lowest cognitive load (red). The dashed lines indicate the means under the respective regimes.

diagnostic orders placed and (ii) hospital admissions. Our simulation suggests that a reallocation based on cognitive load would reduce the number of diagnostic orders by roughly 0.79 (18%) per encounter. Following our earlier results, it is also likely that while reducing the number of overall orders, higher informativeness overall would result from the diagnostic orders placed. For admissions, the reassignment reduces admissions by 7%, highlighting the potential for impacting this important decision via patient reassignment. Further analysis along these lines, with a more developed counterfactual assignment model and some additional assumptions, could assess the impact of other policies, e.g., increases in ED staffing (though accounting for quality differences would be potentially valuable, as it would also be here).

Table 6: Actual Versus Counterfactual Means

	Actual	Counterfactual
Mean number of diagnostic orders	4.552	3.760
Mean inpatient admission rate	0.238	0.222

7 Conclusion

We use novel click-stream data on physician actions in the emergency department to unpack the impacts of cognitive load on medical care. Much of the literature to date on cognitive burden and decision-making studies lab settings or smaller stakes field settings. We extend this literature with an investigation of cognitive load in a high-stakes empirical setting with experts. We observe highly detailed data on case-by-case factors that we can link to beliefs and actions, and also study a setting where there is significant and measurable exogenous variation in cognitive load. We set up a new model of the physician's diagnostic process, incorporating cognitive load and learning, and tie our empirical outcomes to hypotheses generated by the model.

We find that cognitive load impacts care in myriad ways. Holding all else equal for a given patient, higher cognitive load for a physician increases the number of diagnostic orders placed for the patient and shifts orders away from more rare targeted diagnostic orders to more common diagnostic orders. We construct a measure of entropy reduction to assess diagnostic belief refinement, and find that higher cognitive load leads to diagnostic order sets that do a coarser job in refining beliefs about potential diagnoses. We show that a one standard deviation increase in cognitive load is associated with a 9% increase in inpatient hospital admission from the ED, a meaningful change in a very costly and impactful health care decision. Further, we find that providers in the most cognitively loaded decile are 28% more likely to admit a patient then when they are in the least cognitively loaded decile and that those marginally admitted due to cognitive load are healthier (both ex ante and ex post) than the typical admitted patient. We show that our results hold for specific chief complaints (chest pain and abdominal pain) and assess a variety of robustness specifications, including those that change (i) our construction of cognitive load measures and (ii) our definition of key outcome variables.

We see a number of exciting paths forward. While we have a granular characterization of cognitive load and an individual physician's typical behavior, it will be valuable in future work to study physician heterogeneity in depth. Our causal forests approach shows that, in our context, we lack sufficient statistical power to study physician heterogeneity in cognitive load impacts in a rich way. Future work with either (i) a very large sample but similar data depth or (ii) experimental variation in cognitive load combined with deep administrative data can make progress. Heterogeneity in cognitive load impacts could be based on demographic / job characteristics (gender, age, job level, experience), differences in provider practice styles (which can be modeled in detail), and unobserved characteristics (e.g., proneness to stress or mental fatigue).

The external validity of our main findings, and the implications of contextual heterogeneity, is also an important path for future work. We study a large, cutting-edge, emergency department at a prestigious academic medical center. The effects of cognitive load could differ in other contexts, not just because of physician heterogeneity, but also because of factors like hospital staffing, hospital equipment, or hospital business models (with implications for provider incentives). This kind of contextual heterogeneity could potentially allow testing of different cognitive load models if perceived defaults (or best actions with coarser diagnosis distributions) vary by context. A dataset with our depth of data, across myriad contexts, would provide a strong foundation for this analysis.

Our simple counterfactual analysis studies patient reassignment accounting for cognitive load, holding a number of contextual factors fixed. There is room to extend this analysis going forward in a number of directions including, e.g., constructing a more complete dynamic assignment model and then integrating structural estimates of our conceptual diagnosis model as inputs into various policy analyses leveraging that new assignment model. In addition, one could study other labor design policies such as mandated break times, shift staggering, caps on hours worked, or how hospitals trade off number of employees with hours per employee. Additionally, there are important design questions related to team production and how effective teams can mitigate shocks like cognitive load (see, e.g., Chan (2016) and Silver (2020)). Work on team-related questions could leverage staffing shocks or shift-design changes to both identify the implications of teams on cognitive load and answer related design questions.

Finally, there are several directions to expand research on cognitive load in settings like ours. Researches could implement lab-in-the-field or field experiments digging more into specific cognitive load mechanisms in our context, and continue to leverage granular administrative data to enhance our understanding of how cognitive load operates. This could build closely on the recent experimental literature on cognitive load (e.g., Enke and Graeber (2023)) or recent conceptual literature on memory and cognition (e.g., Bordalo et al. (2025)). More broadly, with structural estimates from our conceptual diagnosis model, one could make some progress studying how artificial intelligence tools and/or certain kinds of physician training could improve care in settings with cognitive load. Of course, such studies could have broader impacts than just honing in on cognitive load: it is well known that physicians have different practice styles and that there is still much research to do investigating how altering those practice styles (e.g., to mimic top performing doctors in a granular way) can positively impact productivity (Chandra and Staiger (2007)).

References

Abadie, Alberto, "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 2003, *113* (2), 231–263.

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh, "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care," *American Economic Review*, December 2016, *106* (12), 3730–64.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz, "Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology," Working Paper 31422, National Bureau of Economic Research July 2023.

Arrow, Kenneth J., "Uncertainty and the welfare economics of medical care," *American Economic Review*, December 1963, *53* (5), 941–973.

Athey, Susan, Julie Tibshirani, and Stefan Wager, "Generalized random forests," *The Annals of Statistics*, 2019, *47* (2), 1148 – 1178.

Badinski, Ivan, Amy Finkelstein, Matthew Gentzkow, and Peter Hull, "Geographic Variation in Healthcare Utilization: The Role of Physicians," Working Paper 31749, National Bureau of Economic Research October 2023.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Memory, Attention, and Choice*," *The Quarterly Journal of Economics*, 04 2020, *135* (3), 1399–1442.

_ , _ , Giacomo Lanzani, and Andrei Shleifer, "A Cognitive Theory of Reasoning and Choice," 2025. Harvard working paper.

Caplin, Andrew and Mark Dean, "Revealed preference, rational inattention, and costly information acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

Chan, David, "The Efficiency of Slacking Off: Evidence from the Emergency Department," *Econometrica*, 2018, *86* (3), 997–1030.

Chan, David C., "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, 2016, *124* (3), 734–770.

Chan, David C, Matthew Gentzkow, and Chuan Yu, "Selection with Variation in Diagnostic Skill: Evidence from Radiologists*," *The Quarterly Journal of Economics*, 01 2022, *137* (2), 729–783.

Chandra, Amitabh and Douglas O. Staiger, "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks," *Journal of Political Economy*, 2007, *115* (1), 103–140.

_ , David Cutler, and Zirui Song, "Chapter Six - Who Ordered That? The Economics of Treatment Choices in Medical Care," in Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds., *Handbook of Health Economics*, Vol. 2 of *Handbook of Health Economics*, Elsevier, 2011, pp. 397–432.

Chodick, Gabriel, Yoav Goldstein, Ity Shurtz, and Dan Zeltzer, "Challenging Encounters and within-Physician Practice Variability," *The Review of Economics and Statistics*, 05 2025, pp. 1–9.

Currie, Janet and W. Bentley MacLeod, "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians," *Journal of Labor Economics*, 2017, *35* (1), 1–43.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg, "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," *American Economic Journal: Economic Policy*, February 2019, *11* (1), 192–221.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences*, 2011, *108* (17), 6889–6892.

Dean, Mark and Nathaniel Neligh, "Experimental tests of rational inattention," *Journal of Political Economy*, 2023, *131* (12), 3415–3461.

Doyle, Joseph J., Steven M. Ewer, and Todd H. Wagner, "Returns to physician human capital: Evidence from patients randomized to physician teams," *Journal of Health Economics*, 2010, *29* (6), 866–882.

Enke, Benjamin and Thomas Graeber, "Cognitive Uncertainty*," *The Quarterly Journal of Economics*, 05 2023, *138* (4), 2021–2067.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams, "Sources of Geographic Variation in Health Care: Evidence From Patient Migration*," *The Quarterly Journal of Economics*, 07 2016, *131* (4), 1681–1726.

Gabaix, Xavier, "Chapter 4 - Behavioral inattention," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics - Foundations and Applications 2*, Vol. 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, North-Holland, 2019, pp. 261–343.

Hull, Peter, ""One Weird Trick" to Characterize Effective Populations in Design-Based Specifications," September 2025. Unpublished note, September 3, 2025.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach, "Do financial concerns make workers less productive?," *The Quarterly Journal of Economics*, 2025, *140* (1), 635–689.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions*," *The Quarterly Journal of Economics*, 08 2017, *133* (1), 237–293.

Kloiber, Kevin, "The Cost of Complexity: How Communication Frictions Shape Patient Mortality," 2025. LMU working paper.

Matějka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 2015, *105* (1), 272–298.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," *Journal of Economic Literature*, March 2023, *61* (1), 226–73.

McGuire, Thomas G., "Physician agency," in A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics*, 1 ed., Vol. 1, Elsevier, 2000, chapter 09, pp. 461–536.

Mullainathan, Sendhil and Ziad Obermeyer, "A Machine Learning Approach to Low-Value Health Care: Wasted Tests, Missed Heart Attacks and Mis-Predictions," Working Paper 26168, National Bureau of Economic Research August 2019.

Oprea, Ryan, "Decisions under Risk Are Decisions under Complexity," *American Economic Review*, December 2024, *114* (12), 3789–3811.

Paley, Liza, Taiba Zornitzki, Joel Cohen, Joshua Friedman, Natalia Kozak, and Ami Schattner, "Utility of Clinical Examination in the Diagnosis of Emergency Department Patients Admitted to the Department of Medicine of an Academic Hospital," *Archives of Internal Medicine*, 08 2011, *171* (15), 1393–1400.

Pelaccia, Thierry, Jacques Tardif, Emmanuel Triby, Christine Ammirati, Catherine Bertrand, Valérie Dory, and Bernard Charlin, "How and When Do Expert Emergency Physicians Generate and Evaluate Diagnostic Hypotheses? A Qualitative Study Using Head-Mounted Video Cued-Recall Interviews," *Annals of Emergency Medicine*, 2014, *64* (6), 575–585.

Shanmugam, Priya, "Essays on Decision-Making," 2020. Harvard Doctoral Disseration, Economics Department.

Silver, David, "Haste or Waste? Peer Pressure and Productivity in the Emergency Department," *The Review of Economic Studies*, 08 2020, *88* (3), 1385–1417.

Skinner, Jonathan, "Chapter Two - Causes and Consequences of Regional Variations in Health Care," in Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds., *Handbook of Health Economics*, Vol. 2 of *Handbook of Health Economics*, Elsevier, 2011, pp. 45–93.

Song, Yunjie, Jonathan Skinner, Julie Byn6um, Jason Sutherland, John Wennberg, and Elliot Fisher, "Regional Variations in Diagnostic Practices," *New England Journal of Medicine*, 2010, *363*, 45–53.

UCSF Emergency Department, "Primer 1: Foundational Concepts of Diagnostic Error," 2026. https://codex.ucsf.edu/primer-1-foundational-concepts-diagnostic-error: :text=4,in

Wager, Stefan and Susan Athey, "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

Wellsjo, Alex Steiny, "Simple Actions, Complex Habits: Lessons from Hospital Hand Hygiene," 2025. UCSD working paper.

Xiang, Yang, Benjamin Enke, Thomas Graeber, and Samuel Gershman, "Confidence and Central Tendency in Perceptual Judgement," *Atten Percepts Psychophys.*, 04 2021, *83* (7), 3024–3034.

APPENDIX

A Proofs for Section 2

Proof of Proposition 1.

As expressed in (3), the objective can be decomposed into $U(n) = -(1-2^{-n-3})2^{-n-2} - 2^{-2n-6}$ and the cost of information $C(n^M, n^E, L)$. The health and resource benefit $U()$ depends just on the total number of signals n, while the cost of information also on the split between n^M and n^E.

Thus, let us denote $C^{min}(n, L) = \min_{\{n^M, n^E\}} C(n^M, n^E, L)$ the resulting cost of information as a function of the total number of signals $n = n^M + N^E$ only with the optimal composition of n^M and n^E. Let us also define marginal costs of mental signals, $MC^M(n^M, L) = C^M(n^M, L) - C^M(n^M - 1, L)$, and of external signals $MC^E(n^E) = C^E(n^E) - C^E(n^E - 1)$.

Next, consider a common ordered and increasing sequence of all marginal costs of both types of signals. The total cost $C^{min}(n, L)$ is a sum of marginal costs of the n cheapest signals among all $MC^M(\cdot, L)$ and $MC^E(\cdot, L)$. The marginal cost $MC^{min}(n, L)$ is the cost of the n^{th} cheapest signal - it weakly increasing in n and also in L, since mental costs $MC^M(n^M, L)$ are increasing in L.

Let us now show that the marginal of U, i.e., $MU(n) = U(n) - U(n-1)$, is positive and decreasing in n. Using simple algebra, it equals $2^{-2(n+3)}\left(2^{n+4} - 3\right)$, which is positive for $n \geq 0$. Moreover, $MU(n) - MU(n-1) = -2^{-2(n+3)}\left(2^{n+4} - 9\right)$, which is negative for $n \geq 0$, and thus the marginal $MU(n)$ is decreasing in n.

Since $MU(n)$ is positive and decreasing in n, while $MC^{min}(n, L)$ is positive and increasing in n, the optimal $n^* \geq 0$ is the largest n such that $MU(n) \geq MC^{min}(n, L)$. All signals, regardless of whether mental or external, with a marginal cost lower or equal to $MU(n^*)$ are used.

Now, we are ready to establish the comparative statics. If L increases, then $MC^{min}(n, L)$ weakly increases for all n, and thus the optimal n^* weakly decreases, which is the result (i) of Proposition 1.

This also means that if L increases, then marginal benefit of the last used signal $MU(n^*)$ weakly increases. Since L does not affect costs of external signals, the optimal n^E weakly increases, because the same costs $MC^E()$ are weighed against a higher $MU(n^*)$. This is the result (iii). On the other hand, n^M weakly decreases, since $n^M = n^* - n^E$, the result (ii). Finally, the result (iv) holds as an immediate implication of (i) and of $A = 2^{-n^*-3}$.

□

B Details on Emergency Department Environment

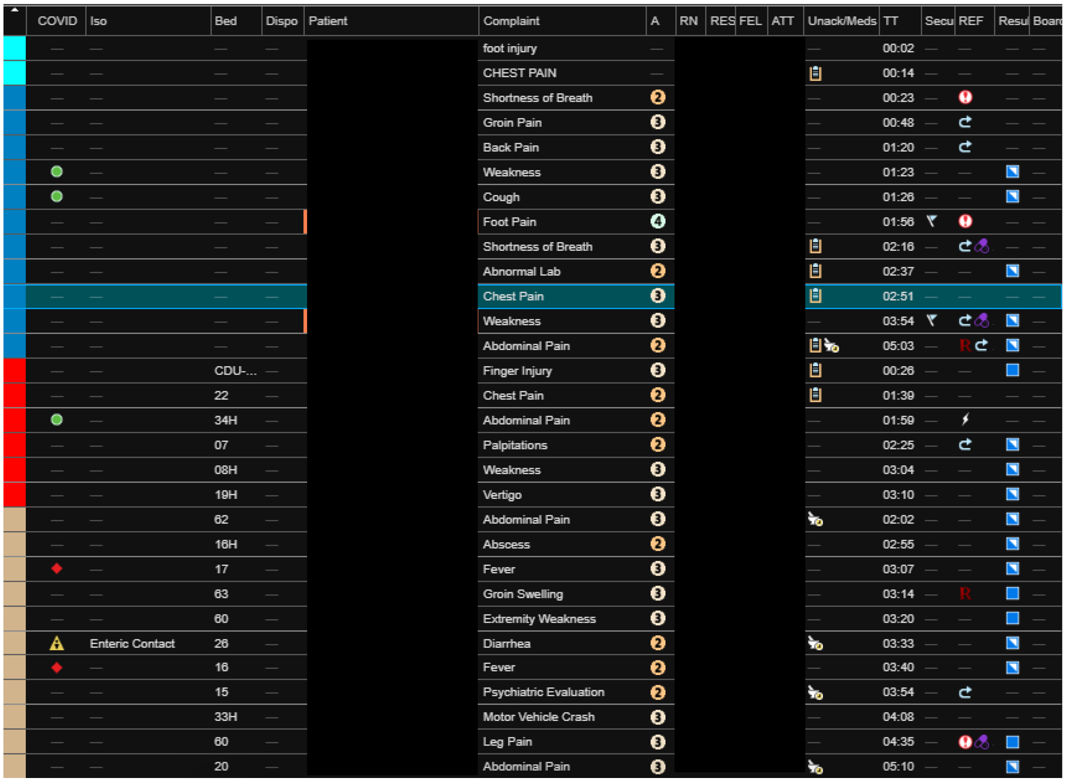
Patients arrive either by ambulance or by walk-in to the front-desk station. A registrar collects demographic identifiers and opens an electronic encounter in Epic. This "rapid registration" usually takes less than 2 minutes and does not influence subsequent triage priority.

A licensed triage nurse then performs a focused assessment in a dedicated triage bay (or at the bedside for direct-to-room arrivals). Vital signs, symptom onset, and brief history are obtained and recorded, together with any required isolation flags. Using the Emergency Severity Index (ESI, version 5) algorithm the nurse assigns an acuity code from 1 (immediate life-saving intervention) to 5 (no resources anticipated). The nurse simultaneously selects a structured chief-complaint label from Epic's menu (e.g., "chest pain", "abdominal pain") or enters free text; this label is the field displayed in Appendix Figure B.1.

Once triage is complete, patients are either taken straight to an available bed ("direct-bed") or returned to the waiting room. Their queue position is managed by the charge nurse but displayed to all clinicians on the live computerized tracking board (shown in Appendix Figure B.1). The board lists, in order of arrival within each acuity tier, the chief complaint, the elapsed waiting time and any preliminary nursing orders such as point-of-care tests.

Physicians pick up cases sequentially ("next-up") from this board. At the moment of selection they observe only the chief complaint and acuity code – as reflected in our control variables.

This workflow means that i) initial assignment of acuity level and chief complaint is nurse-driven and completed before any physician becomes involved, and ii) within each time block the set of patients visible to a given doctor is governed almost entirely by arrival sequence and bed availability rather than by patient characteristics beyond acuity and complaint. These features motivate our quasi-random assignment assumption and the control strategy described in Section 3.1.



COVID	Iso	Bed	Dispo	Patient	Complaint	A	RN	RES	FEL	ATT	Unack/Meds	TT	Secu	REF	Resul	Boar
—	—	—	—	[redacted]	foot injury	—	[redacted]	[redacted]	[redacted]	[redacted]	—	00:02	—	—	—	—
—	—	—	—	[redacted]	CHEST PAIN	—	[redacted]	[redacted]	[redacted]	[redacted]		00:14	—	—	—	—
—	—	—	—	[redacted]	Shortness of Breath	2	[redacted]	[redacted]	[redacted]	[redacted]	—	00:23	—		—	—
—	—	—	—	[redacted]	Groin Pain	3	[redacted]	[redacted]	[redacted]	[redacted]	—	00:48	—		—	—
—	—	—	—	[redacted]	Back Pain	3	[redacted]	[redacted]	[redacted]	[redacted]	—	01:20	—		—	—
	—	—	—	[redacted]	Weakness	3	[redacted]	[redacted]	[redacted]	[redacted]	—	01:23	—	—		—
	—	—	—	[redacted]	Cough	3	[redacted]	[redacted]	[redacted]	[redacted]	—	01:26	—	—		—
—	—	—	—	[redacted]	Foot Pain	4	[redacted]	[redacted]	[redacted]	[redacted]	—	01:56			—	—
—	—	—	—	[redacted]	Shortness of Breath	3	[redacted]	[redacted]	[redacted]	[redacted]		02:16	—		—	—
—	—	—	—	[redacted]	Abnormal Lab	2	[redacted]	[redacted]	[redacted]	[redacted]		02:37	—	—		—
—	—	—	—	[redacted]	Chest Pain	3	[redacted]	[redacted]	[redacted]	[redacted]		02:51	—	—	—	—
—	—	—	—	[redacted]	Weakness	3	[redacted]	[redacted]	[redacted]	[redacted]	—	03:54				—
—	—	—	—	[redacted]	Abdominal Pain	2	[redacted]	[redacted]	[redacted]	[redacted]		05:03	—	R		—
—	—	CDU-...	—	[redacted]	Finger Injury	3	[redacted]	[redacted]	[redacted]	[redacted]		00:26	—	—		—
—	—	22	—	[redacted]	Chest Pain	2	[redacted]	[redacted]	[redacted]	[redacted]		01:39	—	—	—	—
	—	34H	—	[redacted]	Abdominal Pain	2	[redacted]	[redacted]	[redacted]	[redacted]	—	01:59	—		—	—
—	—	07	—	[redacted]	Palpitations	2	[redacted]	[redacted]	[redacted]	[redacted]	—	02:25	—			—
—	—	08H	—	[redacted]	Weakness	3	[redacted]	[redacted]	[redacted]	[redacted]	—	03:04	—	—		—
—	—	19H	—	[redacted]	Vertigo	3	[redacted]	[redacted]	[redacted]	[redacted]	—	03:10	—	—		—
—	—	62	—	[redacted]	Abdominal Pain	3	[redacted]	[redacted]	[redacted]	[redacted]		02:02	—	—		—
—	—	16H	—	[redacted]	Abscess	2	[redacted]	[redacted]	[redacted]	[redacted]	—	02:55	—	—		—
	—	17	—	[redacted]	Fever	3	[redacted]	[redacted]	[redacted]	[redacted]	—	03:07	—	—		—
—	—	63	—	[redacted]	Groin Swelling	3	[redacted]	[redacted]	[redacted]	[redacted]	—	03:14	—	R		—
—	—	60	—	[redacted]	Extremity Weakness	3	[redacted]	[redacted]	[redacted]	[redacted]	—	03:20	—	—		—
	Enteric Contact	26	—	[redacted]	Diarrhea	2	[redacted]	[redacted]	[redacted]	[redacted]		03:33	—	—		—
	—	16	—	[redacted]	Fever	2	[redacted]	[redacted]	[redacted]	[redacted]	—	03:40	—	—		—
—	—	15	—	[redacted]	Psychiatric Evaluation	2	[redacted]	[redacted]	[redacted]	[redacted]		03:54	—		—	—
—	—	33H	—	[redacted]	Motor Vehicle Crash	3	[redacted]	[redacted]	[redacted]	[redacted]	—	04:08	—	—	—	—
—	—	60	—	[redacted]	Leg Pain	3	[redacted]	[redacted]	[redacted]	[redacted]	—	04:35	—			—
—	—	20	—	[redacted]	Abdominal Pain	3	[redacted]	[redacted]	[redacted]	[redacted]		05:10	—	—		—

Figure B.1: Emergency Department Board

Notes: This figure shows the information screen that is available to physicians. It shows all patients who are currently admitted to the emergency department. Names of patients and physicians are removed for data privacy reasons.

C Predicting Patient Complexity

C.1 Constructing the Patient Complexity Score

Our patient-complexity score c_i is designed to proxy the *expected* time and cognitive effort a physician must invest in encounter i. The construction proceeds in four steps.

1. Define three workload targets For every encounter we extract:

i) y_{1i}: **Total number of diagnostic and therapeutic orders** placed in the ED (medications, labs, imaging, procedures,...)

ii) y_{2i}: **Total note-edit time** in seconds, transformed as $\log(1 + y_{2i})$ to down-weight the right tail.

iii) y_{3i}: **Number of specialty consultations** initiated in the ED.

2. Assemble encounter-level predictors $\mathbf{X}_i$ All predictors are observable at, or immediately after, triage:

- *Demographics and comorbidity:* age-bin dummies ($18–29, 30–44, \ldots, \geq 90$), gender indicator, Charlson comorbidity index (continuous), and an indicator for interpreter need.
- *Chief complaint:* three mutually exclusive sets of dummies – Tier 1 (15 body-system buckets), Tier 2 (72 sub-systems), and the detailed chief-complaint code (732 categories).
- *Clinical acuity and vitals:* ESI acuity code (1–5), six binary "abnormal vital" flags (systolic/diastolic BP, pulse, respiration, temperature, O_2 saturation).
- *Utilization history:* indicators for any ED visit or inpatient admission in the preceding 60 days and for means of arrival (ambulance, police, walk-in, ...)
- *Fixed effects:* 24 hour-of-day and 7 day-of-week dummies.

Continuous variables are z-scored, categorical variables are one-hot-encoded with the first level omitted.

3. Predict each workload target with LASSO regression For $m \in \{1, 2, 3\}$ we regress y_{mi} on $\mathbf{X}_i$ using a five-fold cross-validated LASSO:

$$\hat{\beta}_m \;=\; \arg\min_{\beta} \; \frac{1}{N} \sum_i (y_{mi} - \mathbf{X}_i'\beta)^2 + \lambda \|\beta\|_1, \tag{13}$$

where the penalty λ is chosen by the minimum average cross-validated MSE (Appendix Figure C.1). We then subtract the fitted contributions of the hour- and day-dummies so that the resulting predicted values, $\hat{y}_{mi}^{\text{FE-adj}}$, vary only with patient covariates.

4. Create a single complexity index Each FE-adjusted prediction is standardized:

$$\tilde{y}_{mi} = (\hat{y}_{mi}^{\text{FE-adj}} - \mu_m)/\sigma_m \tag{14}$$

The composite workload proxy is their simple average:

$$\tilde{c}_i = \tfrac{1}{3} \sum_{m=1}^{3} \tilde{y}_{mi}. \tag{15}$$

We standardize once more and shift by the absolute minimum so that

$$c_i \;=\; \frac{\tilde{c}_i - \mu_c}{\sigma_c} \;+\; \left|\min_j \tfrac{\tilde{c}_j - \mu_c}{\sigma_c}\right| \quad >\; 0. \tag{16}$$

The distribution is right-skewed with a primary mode near 3 (Appendix Figure D.3).

C.2 Alternative LASSO Specifications

We estimate six variants of the prediction exercise to verify the robustness of c_i:

Chief-complaint covariates	Full sample	Two-half split
Tier 2 only	✓	✓
Detailed code only	✓	✓
Tier 1 + Tier 2 + detailed	✓	✓

Two-half sample split: To eliminate any possibility of over-fitting "leaking" into the main analysis, we randomly assign encounters to halves A and B. A model trained on A predicts complexity for B and *vice-versa*. Each observation is therefore scored by a model that never used its own outcome data.

Chief-complaint granularity: Using only Tier 2, only the detailed code, or all three levels in the design matrix tests whether the predictive content of chief complaint depends on resolution.

Across all six variants the within-sample R^2 values differ by < 0.01 (Appendix Table C.4), and the pairwise correlations of the resulting complexity indices exceed 0.96 (same table). We adopt the *full-sample*, *Tier1 + Tier2 + detailed* specification for the baseline results for the sake of simplicity.

C.3 Summary Statistics of Predictors

The most important predictors for the number of orders placed for a given encounter are non-urgent and less-urgent acuity assessments which both enter negatively. Chief complaints related to the endocrine system positively predict the number of orders, likely since this tier comprises an array of relatively distinct conditions (e.g., diabetes, thyroid disorders, hormonal imbalances). Non-urgent or missing acuity codes also substantially decrease the predicted note edit time for the encounter while urgent or emergent acuity assessments increase the predicted time spent editing notes. Chief complaints in the mental health category ("psychiatric" or "suicidal") lead to more consultations. Across all three variables, the LASSO coefficients always agree in their sign and are mostly extremely similar in magnitude.

Table C.1: Summary Statistics for LASSO Outcomes and Continuous Predictors

	Min	P10	P25	Mean	Median	P75	P90	Max
Outcomes								
Number of all orders in ED	0	1	4	16.737	11	23	41	209
Number of consults	0	0	0	0.232	0	0	1	13
Log(1 + Total ED note edit time)	0.000	5.787	6.268	6.809	6.834	7.515	8.118	10.041
Continuous predictors								
Charlson comorbidity index	0	0	0	0.617	0	0	2	16
Number of abnormal vital signs	0	0	0	0.565	0	1	2	6

Notes: The table shows summary statistics for the three outcomes and the two continuous predictors of the LASSO complexity prediction model.

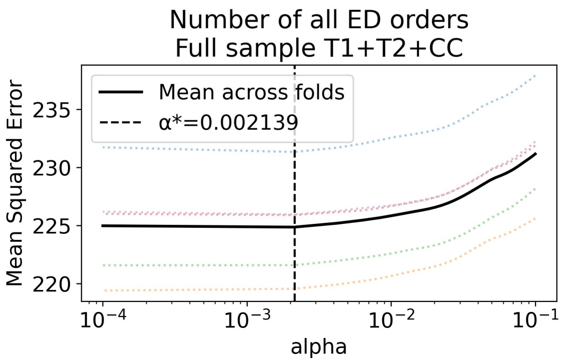
Table C.2: Summary Statistics for Categorical LASSO Predictors

	Share
Demographics	
Gender: Male	0.510
Interpreter needed	0.102
Health status	
Admitted within last 60 days	0.432
ED visit within last 60 days	0.217
Vitals	
Abnormal systolic blood pressure	0.099
Abnormal diastolic blood pressure	0.316
Abnormal temperature	0.004
Abnormal pulse oximetry	0.067
Abnormal Respiration	0.052
Abnormal pulse	0.028
Acuity level	
Immediate	0.004
Emergent	0.194
Urgent	0.599
Less Urgent	0.180
Non-Urgent	0.019
Missing	0.003
Age bin	
18-29	0.197
30-39	0.162
40-49	0.142
50-59	0.169
60-69	0.160
70-79	0.106
80+	0.064
N = 84,214	

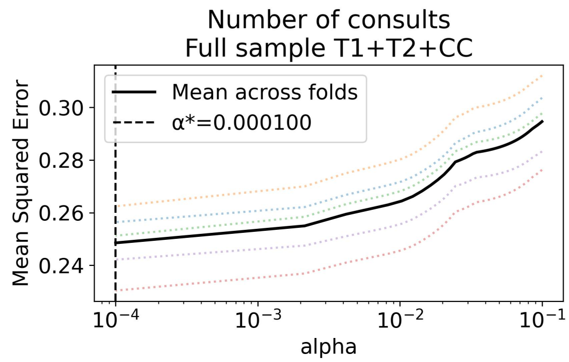
	Share
Chief complaint Tier 1 (15 total categories)	
Neurological	0.147
GI/Abdominal	0.143
Musculoskeletal	0.135
Other	0.099
Cardiovascular	0.085
Respiratory	0.072
Trauma	0.068
HEENT	0.049
Psychiatric	0.049
GU	0.045
Chief complaint Tier 2 (72 total categories)	
MSK pain	0.110
Abdominal pain	0.098
Chest pain	0.062
Respiratory problem	0.057
Neurological symptoms	0.038
Psychiatric problem	0.037
Vomiting/diarrhea	0.030
Fall	0.029
Headache	0.028
Altered Mental Status	0.027
Chief complaint (732 total categories)	
Abdominal Pain	0.096
Chest Pain	0.062
Shortness of Breath	0.053
Fall	0.029
Altered Mental Status	0.027
Headache	0.026
Back Pain	0.025
Dizziness	0.023
Psychiatric Evaluation	0.022
Weakness	0.022

Notes: The table shows the category shares of all categorical predictors used in the LASSO model to predict patient complexity. For the chief complaint variables (Tier 1, Tier 2, and detailed), only the ten most common categories are shown.

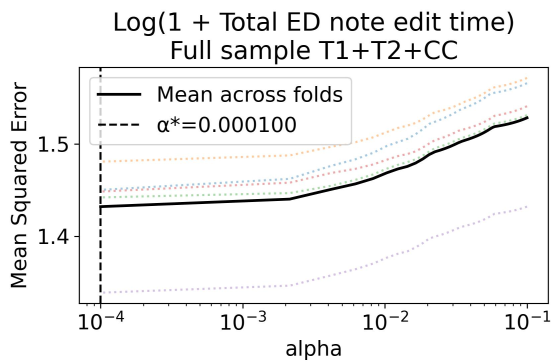
C.4 Evaluating Lasso Predictions



Panel A: Number of all ED orders



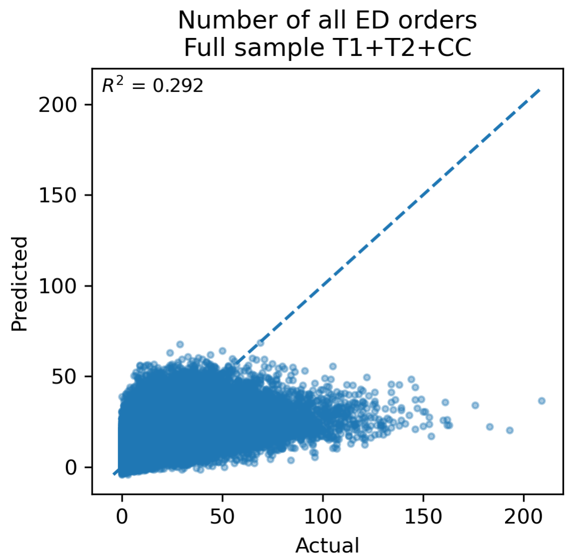
Panel B: Number of consults



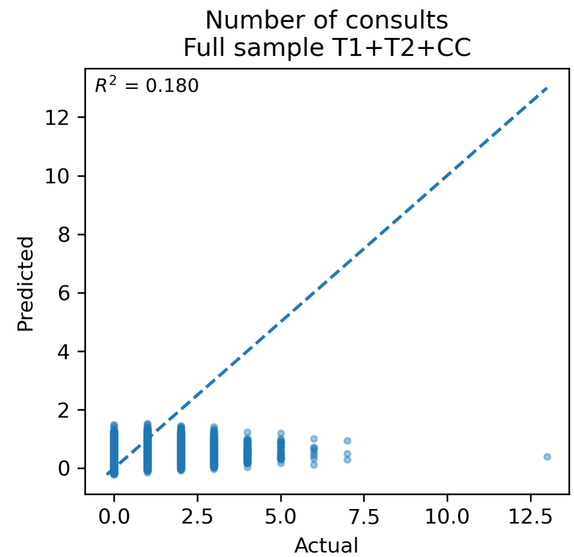
Panel C: Log(1 + Total ED note edit time)

Figure C.1: Five-fold LASSO cross-validation results

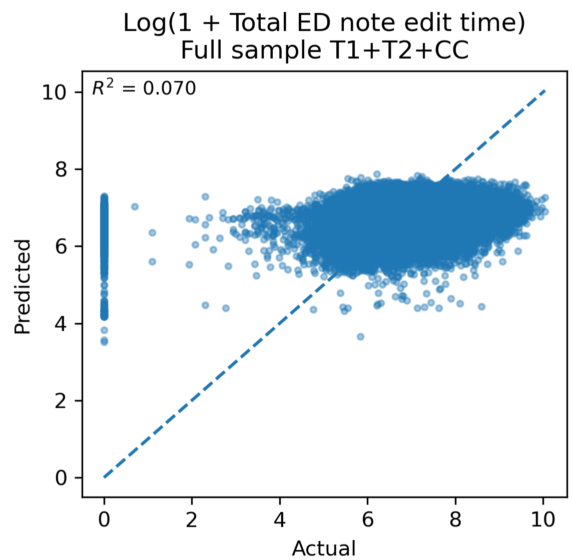
Notes: All results are for the full-sample specification using all three chief complaint categorical variables as predictors: Tier 1 (T1), Tier 2 (T2), and detailed (CC) chief complaint. Dotted lines show fold-specific mean-squared-error (MSE) trajectories across the tested regularization parameter (α) grid; the solid black line is the average across folds. The vertical dashed line marks the α value chosen by cross-validation.



Panel A: Number of all ED orders



Panel B: Number of consults



Panel C: Log(1 + Total ED note edit time)

Figure C.2: Comparison of model predictions to observed values.

Notes: All results are for the full-sample specification using all three chief complaint categorical variables as predictors: Tier 1 (T1), Tier 2 (T2), and detailed (CC) chief complaint. Each panel plots encounter-level predictions (vertical axis) against actual outcomes (horizontal axis); the 45° dashed line indicates perfect prediction. The within-sample R^2 for each outcome is reported in the upper-left corner of its panel.

C.5 Comparing Different Lasso Prediction Models

Table C.3: Goodness-of-Fit (R^2) for Patient Complexity Measures

	Number of all ED orders	Log(Note edit time)	Number of consults
Full sample T1 + T2 + CC	0.292	0.070	0.180
Sample split T1 + T2 + CC	0.290	0.068	0.172
Full sample CC	0.292	0.069	0.180
Sample split CC	0.290	0.067	0.177
Full sample T2	0.286	0.066	0.164
Sample split T2	0.285	0.064	0.163

Notes: Entries show within-sample R^2 for each LASSO model. "Sample split" fits on one half of the data and is evaluated on the other; "Full sample" uses the entire dataset. Chief-complaint detail enters the LASSO as Tier 2 indicators ("T2"), the detailed chief-complaint code ("CC"), or all three levels combined ("T1 + T2 + CC").

Table C.4: Correlation Matrix for Patient Complexity Measures

	T2 (split)	T2 (full)	CC (split)	CC (full)	T1+T2+CC (split)	T1+T2+CC (full)
T2 (split)	1.000	0.998	0.983	0.982	0.989	0.984
T2 (full)	0.998	1.000	0.982	0.984	0.987	0.986
CC (split)	0.983	0.982	1.000	0.996	0.995	0.995
CC (full)	0.982	0.984	0.996	1.000	0.993	0.999
T1 + T2 + CC (split)	0.989	0.987	0.995	0.993	1.000	0.994
T1 + T2 + CC (full)	0.984	0.986	0.995	0.999	0.994	1.000

Notes: Each entry shows the Pearson correlation between two encounter-level complexity indices. All indices are standardized, shifted versions of the underlying composite score described in Section 2.3. "(split)" refers to "Sample split": these models are fit on one half of the data and evaluated on the other; "(full)" refers to "Full sample": these models are estimated on the entire dataset. Chief-complaint information enters the LASSO as either Tier 2 indicators ("T2"), the detailed chief-complaint code ("CC"), or all three levels combined ("T1 + T2 + CC").

Table C.5: LASSO Feature Importance: Number of all ED orders

Variable	Coefficient
Chief complaint: Pneumonia	12.469
Acuity code: Non-Urgent	-12.257
Chief complaint: Diabetic Ketoacidosis	11.181
Chief complaint: Diplopia	11.072
Acuity code: Less Urgent	-10.681
Chief complaint: Multiple Sclerosis	9.984
Chief complaint Tier 1: Endocrine	9.472
Chief complaint: Sickle Cell Pain Crisis	8.860
Chief complaint: Meningitis	8.220
Chief complaint: Spasms	8.006
Chief complaint: Hand Problem	6.880
Chief complaint: Arm Swelling	6.779
Chief complaint: Gait Problem	6.694
Chief complaint: Hallucinations	-6.347
Chief complaint: Hypertension	-6.250
Chief complaint: Extremity Weakness	5.749
Acuity code: Missing	-5.500
Means of arrival: Transfer - System Arrival	-5.103
Chief complaint Tier 1: Constitutional	4.728
Chief complaint Tier 2: Chest pain	4.364
Chief complaint: Brain Tumor	-4.241
Chief complaint: Cellulitis/ Skin Infection	4.234
Chief complaint: Sciatica	-4.233
Chief complaint: Shortness of Breath	4.192
Chief complaint Tier 2: Flank pain	4.172

Notes: This table shows the 25 most important predictors for the outcome "Number of all ED orders" as measured by the absolute magnitude of the LASSO coefficient.

Table C.6: LASSO Feature Importance: Log(1 + Total ED note edit time)

Variable	Coefficient
Chief complaint: Other	-1.015
Acuity code: Missing	-1.004
Means of arrival: Unknown	-0.823
Acuity code: Non-Urgent	-0.579
Chief complaint Tier 2: Social problems	-0.414
Chief complaint: Drug / Alcohol Assessment	-0.327
Chief complaint: Alcohol Intoxication	-0.282
Chief complaint: Hallucinations	-0.277
Chief complaint: Asthma	-0.243
Acuity code: Emergent	0.225
Chief complaint Tier 1: Drug use	-0.211
Chief complaint: Leg Injury	-0.197
Chief complaint Tier 2: Psychiatric problem	-0.184
ED visit within last 60 days: 1	-0.170
Acuity code: Urgent	0.166
Age bin: 30-39	-0.164
Acuity code: Less Urgent	-0.163
Chief complaint: Shoulder Injury	-0.160
Chief complaint Tier 2: Blood disorders	0.156
Chief complaint Tier 2: Weight loss	0.136
Chief complaint: Rash	0.135
Chief complaint: Psychiatric Evaluation	0.134
Age bin: 40-49	-0.129
Chief complaint: Cerebrovascular Accident	-0.125
Age bin: 18-29	-0.124

Notes: This table shows the 25 most important predictors for the outcome "Log(1 + Total ED note edit time)" as measured by the absolute magnitude of the LASSO coefficient.

Table C.7: LASSO Feature Importance: Number of Consults

Variable	Coefficient
Chief complaint: Multiple Sclerosis	0.590
Chief complaint Tier 2: Suicidal	0.532
Chief complaint: Airway Obstruction	0.481
Chief complaint: Urinary Incontinence	0.442
Chief complaint: Diplopia	0.433
Chief complaint: Zoster keratitis	0.430
Chief complaint: Circulatory Problem	0.421
Chief complaint: Manic Behavior	0.384
Chief complaint: Depression	0.369
Chief complaint: Psychiatric Evaluation	0.355
Chief complaint: Aggressive Behavior	0.352
Chief complaint: Foot Ulcer	0.344
Chief complaint: Skin Lesion	0.330
Chief complaint: Shingles/ Herpes Zoster	0.284
Chief complaint: Sciatica	-0.274
Chief complaint: Loss of Vision	0.270
Chief complaint: Gait Problem	0.266
Chief complaint: Stress	-0.266
Chief complaint: Floaters	0.255
Chief complaint: GI Problem	0.251
Chief complaint: Anal Pain	0.247
Means of arrival: Police	0.240
Acuity code: Non-Urgent	-0.231
Chief complaint: Panic Attack	-0.228
Chief complaint Tier 2: Eye problem	0.227

Notes: This table shows the 25 most important predictors for the outcome “Number of consults” as measured by the absolute magnitude of the LASSO coefficient.

D Additional Descriptive Figures and Tables

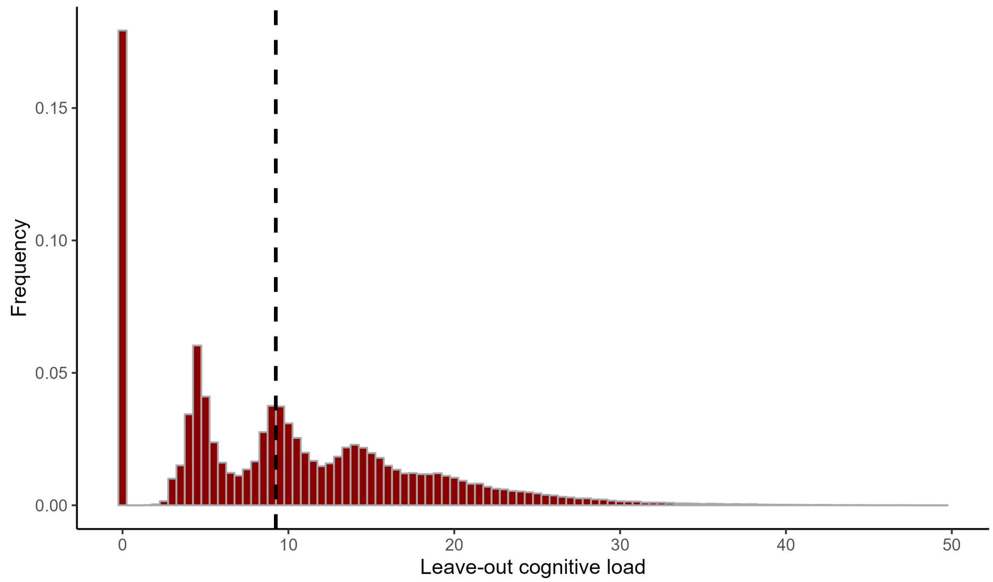


Figure D.1: Distribution of Leave-Out Cognitive Load

Notes: The figure shows the relative frequencies of the leave-out cognitive load measure with a binwidth of 0.5. Each observation is the cognitive load at one *encounter-action*. The vertical dashed line indicates the median.

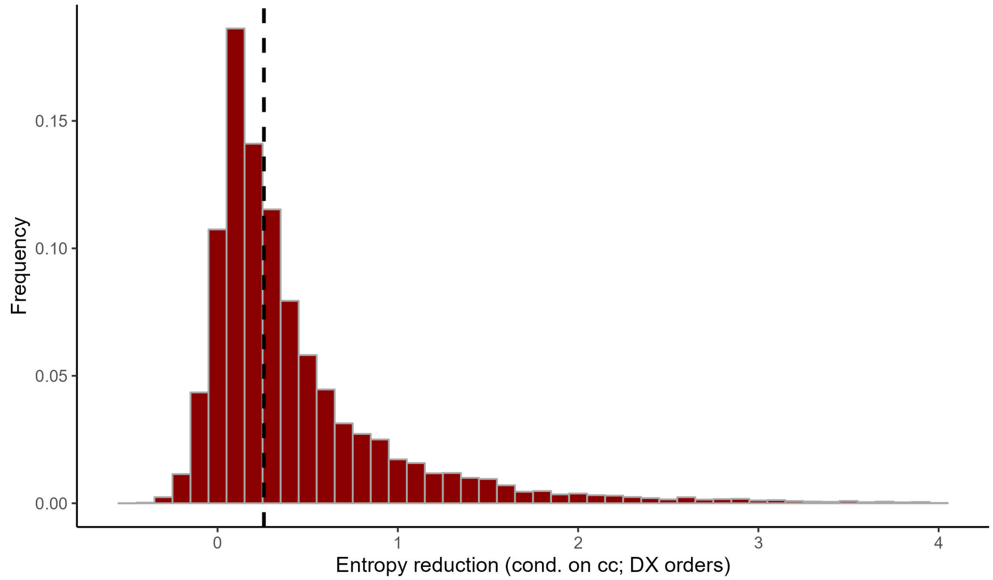


Figure D.2: Distribution of Entropy Reduction in First Order Batch

Notes: The figure shows the relative frequencies of the entropy reduction in the first-order batch with a binwidth of 0.1. Each observation is the entropy reduction for one *encounter*. The vertical dashed line indicates the median.

Table D.1: Shift summary statistics by physician type

	Min	Mean	Median	Max
Residents years 1/2				
Number of shifts	1	20.279	12	142
Average shift length (hours)	7.305	11.786	10.478	28.801
Average patients per shift	1.000	3.505	3.467	9.571
Residents years 3+				
Number of shifts	1	15.642	2	101
Average shift length (hours)	5.165	13.418	12.258	29.178
Average patients per shift	1.000	3.236	1.000	11.889
Attending physicians				
Number of shifts	1	71.897	13	514
Average shift length (hours)	5.387	11.270	10.734	24.794
Average patients per shift	1.000	1.479	1.217	3.431
Nurse practitioner				
Number of shifts	1	92.733	54	412
Average shift length (hours)	7.251	11.137	10.899	15.955
Average patients per shift	1.000	3.487	3.437	8.379

Notes: Each row reports summary statistics for shifts at the *physician* level for the indicated provider category.

Table D.2: Most Frequently Placed Diagnostic Orders

	Order	Share
1	Complete Blood Count With Differential	0.0925
2	Basic Metabolic Panel (Na, K, Cl, Co2, Bun, Cr, Glu, Ca)	0.0594
3	Venous Blood Gas W/Lactate	0.0467
4	Troponin I	0.0416
5	Comprehensive Metabolic Panel (Bmp, Ast, Alt, T.bili, Alkp, Tp Alb)	0.0356
6	Urinalysis With Reflex To Culture	0.0350
7	Urine Culture From Screening Urinalysis	0.0350
8	Peripheral Blood Culture	0.0324
9	Prothrombin Time	0.0299
10	Xr Chest 2 Views Pa And Lateral	0.0293
11	Lipase	0.0250
12	Magnesium, Serum / Plasma	0.0227
13	Bilirubin, Total	0.0208
14	Alanine Transaminase	0.0208
15	Aspartate Transaminase	0.0208
16	Alkaline Phosphatase	0.0207
17	Xr Chest 1 View Ap	0.0164
18	B-Type Natriuretic Peptide	0.0146
19	Poct Urine Pregnancy, ¿= 18 Years	0.0141
20	Ct Brain Without Contrast	0.0140
21	Poct Glucose	0.0137
22	Phosphorus, Serum / Plasma	0.0131
23	Ct Abdomen/Pelvis With Contrast	0.0113
24	Drugs Of Abuse Screen, Rapid	0.0108
25	Type And Screen	0.0105

Notes: The table shows the 'display names' of the 25 most frequently placed diagnostic ED orders, ranked by their share of all diagnostic orders.

Table D.3: Most frequently placed medication orders

	Order	Share
1	Sodium Chloride 0.9 % Iv Bolus	0.0850
2	Sodium Chloride 0.9 % (Flush) Injection Syringe	0.0561
3	Ondansetron Hcl (Pf) 4 Mg/2 Ml Injection Solution	0.0550
4	Acetaminophen 500 Mg Tablet	0.0395
5	Acetaminophen 325 Mg Tablet	0.0308
6	Ondansetron Hcl 4 Mg Tablet	0.0205
7	Ibuprofen 600 Mg Tablet	0.0190
8	Fentanyl (Pf) 50 Mcg/Ml Injection Solution	0.0165
9	Oxycodone 5 Mg Tablet	0.0150
10	Electrolyte-A Iv Bolus	0.0149
11	Lorazepam 2 Mg/Ml Injection Solution	0.0132
12	Lorazepam 1 Mg Tablet	0.0126
13	Potassium Chloride 10 Meq/100ml In Sterile Water Intravenous Piggyback	0.0122
14	Morphine 4 Mg/Ml Intravenous Solution	0.0110
15	Hydromorphone (Pf) 0.5 Mg/0.5 Ml Injection Syringe	0.0104
16	Ipratropium Bromide 0.02 % Solution For Inhalation	0.0101
17	Ketorolac 15 Mg/Ml Injection Solution	0.0099
18	Potassium Chloride Er 20 Meq Tablet,Extended Release(Part/Cryst)	0.0094
19	Vancomycin 1 G Ivpb (Vial-Mate)	0.0090
20	Acetaminophen 1,000 Mg/100 Ml (10 Mg/Ml) Intravenous Solution	0.0089
21	Hydrocodone 5 Mg-Acetaminophen 325 Mg Tablet	0.0089
22	Sodium Chloride 0.9 % Intravenous Solution	0.0084
23	Ceftriaxone 1 G Iv 50 Ml (Mini-Bag Plus)	0.0082
24	Lidocaine (Pf) 10 Mg/Ml (1 %) Injection Solution	0.0082
25	Diphenhydramine 50 Mg/Ml Injection Solution	0.0080

Notes: The table shows the 'display names' of the 25 most frequently placed medication ED orders, ranked by their share of all medication orders.

Table D.4: Most common chief complaints

	Chief complaint	Share
1	Abdominal Pain	0.1062
2	Chest Pain	0.0675
3	Shortness Of Breath	0.0591
4	Fall	0.0302
5	Altered Mental Status	0.0289
6	Headache	0.0261
7	Dizziness	0.0250
8	Weakness	0.0244
9	Fever	0.0241
10	Psychiatric Evaluation	0.0238
11	Back Pain	0.0236
12	Emesis	0.0181
13	Leg Pain	0.0167
14	Other	0.0165
15	Syncope	0.0158
16	Flank Pain	0.0151
17	Seizures	0.0145
18	Cough	0.0140
19	Alcohol Intoxication	0.0130
20	Suicidal	0.0128
21	Knee Pain	0.0093
22	Palpitations	0.0091
23	Foot Pain	0.0089
24	Abnormal Lab	0.0085
25	Leg Swelling	0.0081

Notes: The table shows the 25 most common chief complaints (out of 732 distinct chief complaints).

Table D.5: Most common primary diagnoses

	Diagnosis	Share
1	Chest Pain, Unspecified	0.0356
2	Unspecified Abdominal Pain	0.0301
3	Sepsis, Unspecified Organism (Cms Code)	0.0255
4	Other Chest Pain	0.0190
5	Syncope And Collapse	0.0144
6	Dizziness And Giddiness	0.0139
7	Headache	0.0119
8	Urinary Tract Infection, Site Not Specified	0.0104
9	Epigastric Pain	0.0098
10	Pneumonia, Unspecified Organism	0.0096
11	Unspecified Convulsions (Cms Code)	0.0076
12	Palpitations	0.0073
13	Shortness Of Breath	0.0073
14	Unspecified Fall, Initial Encounter	0.0071
15	Low Back Pain	0.0071
16	Chronic Obstructive Pulmonary Disease With (Acute) Exacerbation	0.0070
17	Calculus Of Kidney	0.0068
18	Nausea With Vomiting, Unspecified	0.0068
19	Fever, Unspecified	0.0067
20	Abnormal Electrocardiogram (Ecg) (Ekg)	0.0059
21	Hypertensive Heart And Chronic Kidney Disease With Heart Failure	0.0056
22	Altered Mental Status, Unspecified	0.0055
23	Fall On Same Level From Slipping, Tripping And Stumbling	0.0053
24	Major Depressive Disorder, Single Episode, Unspecified	0.0053
25	Right Lower Quadrant Pain	0.0053

Notes: The table shows the 25 most common diagnoses.

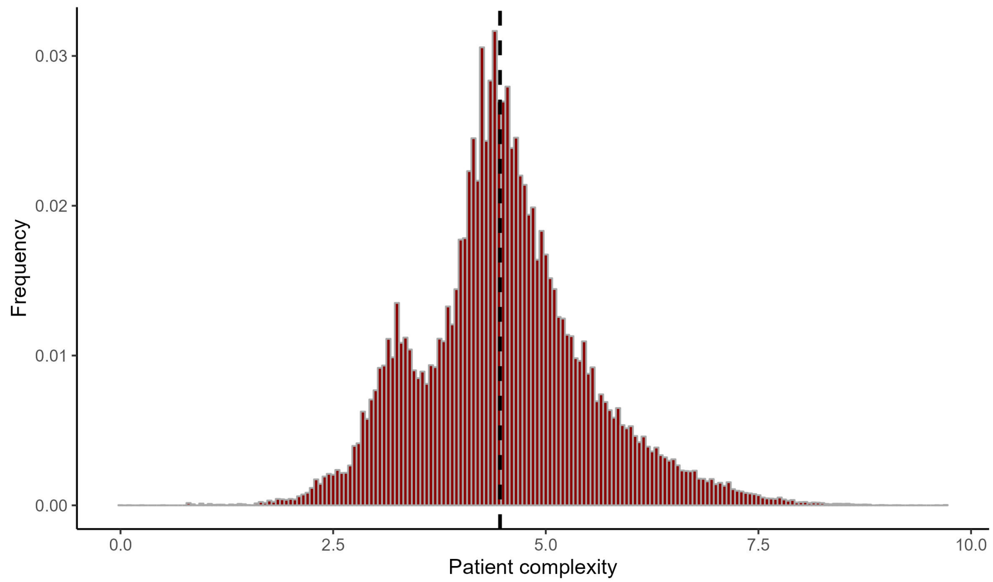


Figure D.3: Distribution of Patient Complexity Measure

Notes: The figure shows the distribution of the patient complexity measure for the full-sample specification using all three chief complaint categorical variables as predictors: Tier 1 (T1), Tier 2 (T2), and detailed (CC) chief complaint. The vertical dashed line shows the median.

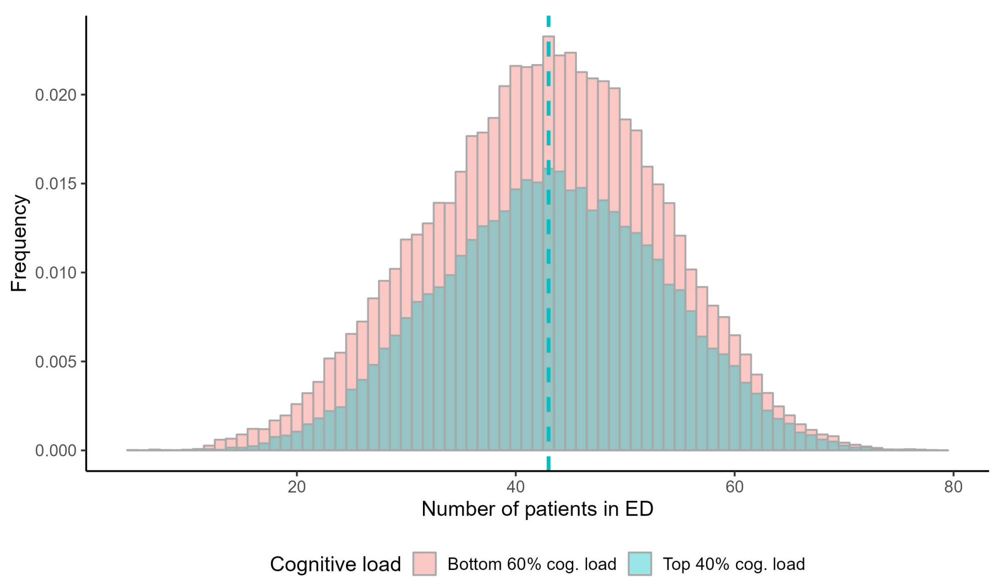


Figure D.4: Distribution of Number of Patients in ED

Notes: The figure shows the distribution of the number of patients in the ED, separately for the top two quintile (blue) and the bottom three quintiles (red). The vertical dashed line shows the medians of both distributions, which overlap exactly.

E Patient-Level Examples of Diagnostic Entropy Reduction

This section illustrates how entropy reduction varies across two abdominal-pain encounters, detailing the diagnostic orders and the calculation procedure introduced in Section 6.1 of the main text.

Patient Example 1

This example follows the entropy-reduction procedure introduced in Section 6.1 and Equations **??**–12 of the main text.

Clinical context: An adult patient presents with *abdominal pain*. Four point-of-care (POC) or laboratory urine studies are ordered within 30 minutes. The encounter's final diagnosis is *urinary-tract infection* (UTI).

Order 1: *POC Urinalysis Dipstick.* Screens urine for glucose, protein, blood, leukocyte esterase, nitrites, ketones and more, offering an immediate infection- or renal-injury signal.

Order 2: *POC Urine Pregnancy Test.* Detects human chorionic gonadotropin (hCG) in minutes, ruling out pregnancy before imaging or medication that is teratogenic.

Order 3: *Laboratory Urinalysis with Reflex-to-Culture.* Microscopic sediment exam; automatically cultures if pyuria or bacteriuria exceeds preset thresholds.

Order 4: *Urine Culture from Screening Urinalysis.* Provides definitive pathogen identification and antimicrobial sensitivities for targeted therapy.

Order-specific diagnosis distributions: Figure E.1 displays four separate histograms - one per order - showing how often each final diagnosis occurs in encounters where that order was placed for abdominal-pain presentations.

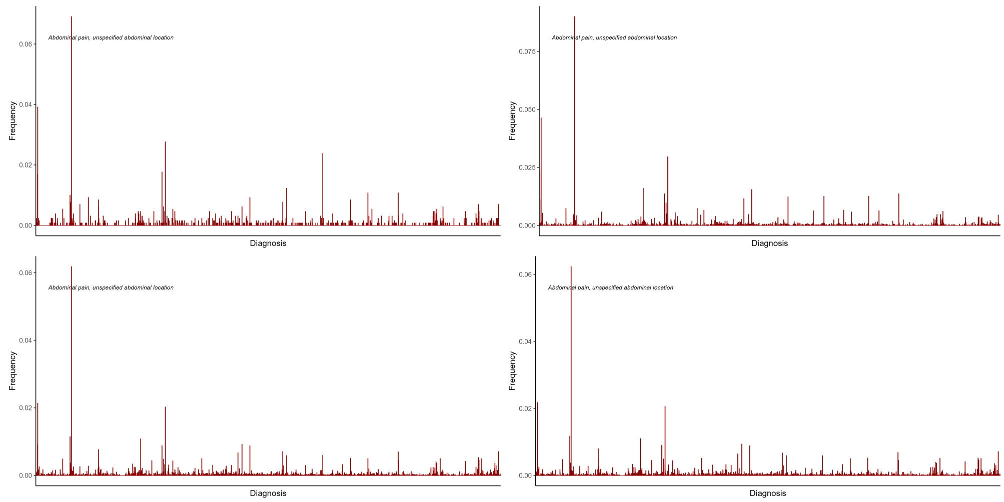


Figure E.1: Diagnosis Distributions Implied by Orders 1–4

Notes: Each panel shows the empirical frequency of final diagnoses associated with an *individual* order, conditional on chief complaint *abdominal pain*. The vertical line marks the diagnosis most frequently linked to that specific order.

Aggregated diagnosis distribution: Figure E.2 *sums* the frequencies from Figure E.1 across all four orders, normalizes them to sum to one, and visualizes the aggregate probability mass function implied by the entire order set.

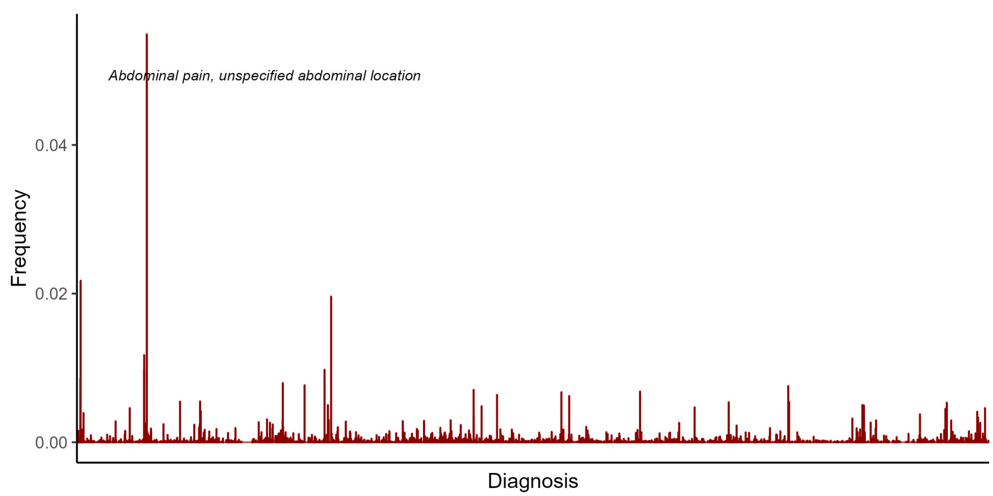


Figure E.2: Aggregated Diagnosis Distribution Across Orders 1–4

Notes: Empirical frequencies from Figure E.1 are summed across the four orders and renormalized to sum to one. The resulting distribution reflects the combined diagnostic information of the full order set. The vertical line indicates the diagnosis with the highest aggregated probability.

Entropy calculation:

1. The empirical entropy of final diagnoses for the chief complaint *abdominal pain* is

$$H(\text{Abdominal Pain}) = 6.52.$$

   A higher entropy value indicates a broader, more uncertain distribution over possible diagnoses.

2. Aggregating the marginal diagnosis distributions implied by Orders 1–4 and calculating the entropy following Equation **??** yields:

$$H(\text{Order Set} \mid \text{Abdominal Pain}) = 5.82.$$

   A *lower* entropy here reflects that the order set places probability mass on a broader subset of diagnoses.

3. Following Equation 12, entropy reduction for this encounter is

$$E_1 = H(\text{Abdominal Pain}) - H(\text{Order Set} \mid \text{Abdominal Pain}) = 0.70.$$

   A *larger* entropy reduction signifies that the physician's diagnostic orders have sharpened the implied distribution which indicates more precise diagnostic reasoning for this patient.

Patient Example 2

Clinical context: A second adult patient presents with *abdominal pain*. Four diagnostic studies—two urine-based and two blood-based—are ordered within 30 minutes. The encounter's final diagnosis is again *urinary-tract infection* (UTI).

Order 1: *Urinalysis with Microscopy.* Full dipstick plus microscopic sediment examination (cells, casts, crystals) performed up-front.

Order 2: *Complete Blood Count with Differential.* Evaluates leukocytosis or left shift that could suggest systemic infection or alternate intra-abdominal pathology.

Order 3: *Basic Metabolic Panel.* Measures electrolytes, renal function and acid–base status, guiding imaging contrast use or antibiotic dosing.

Order 4: *Venous Blood Gas with Lactate.* Provides pH, pCO_2, bicarbonate and lactate; elevated lactate flags sepsis or tissue hypoperfusion.

Order-specific diagnosis distributions: Figure E.3 displays four separate histograms – one per order – showing how often each final diagnosis occurs in encounters where that order was placed for abdominal-pain presentations.

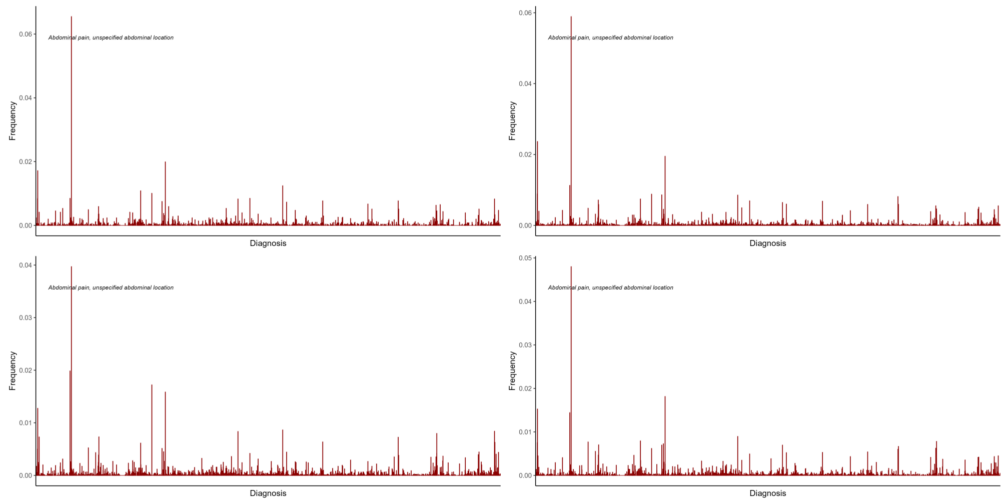


Figure E.3: Diagnosis Distributions Implied by Orders 1–4

Notes: Each panel shows the empirical frequency of final diagnoses associated with an *individual* order, conditional on chief complaint *abdominal pain*. The vertical line marks the diagnosis most frequently linked to that specific order.

Aggregated diagnosis distribution: Figure E.4 *sums* the frequencies from Figure E.3 across all four orders, normalizes them to sum to one, and visualizes the aggregate probability mass function implied by the entire order set.

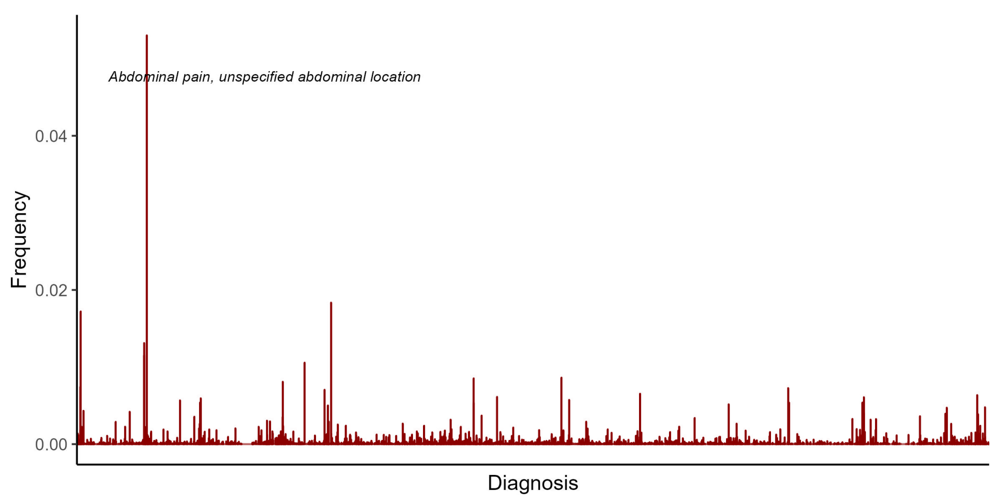


Figure E.4: Aggregated Diagnosis Distribution Across Orders 1–4

Notes: Empirical frequencies from Figure E.3 are summed across the four orders and re-normalized to sum to one. The resulting distribution reflects the combined diagnostic information of the full order set. The vertical line indicates the diagnosis with the highest aggregated probability.

Entropy calculation:

1. The empirical entropy of final diagnoses for the chief complaint *abdominal pain* is

$$H(\text{Abdominal Pain}) = 6.52.$$

   A higher entropy value indicates a broader, more uncertain distribution over possible diagnoses.

2. Aggregating the marginal diagnosis distributions implied by Orders 1–4 and calculating the entropy following Equation **??** yields:

$$H(\text{Order Set} \mid \text{Abdominal Pain}) = 6.44.$$

   A *lower* entropy here reflects that the order set places probability mass on a broader subset of diagnoses.

3. Following Equation 12, entropy reduction for this encounter is

$$E_2 = H(\text{Abdominal Pain}) - H(\text{Order Set} \mid \text{Abdominal Pain}) = 0.08.$$

   A *larger* entropy reduction signifies that the physician's diagnostic orders have sharpened the implied distribution which indicates more precise diagnostic reasoning for this patient. Conversely, the relatively small reduction here suggests a broader, less discriminating work-up.

F Alternative Regression Specifications: Encounter-Action Level

F.1 Alternative Entropy Reduction Specifications

Table F.1: OLS Estimates of the Impact of Cognitive Load on Diagnostic Entropy Reduction

	Within 20 min of First Order	Within 30 min of First Order	Within 60 min of First Order
Panel A: Diagnosis ID			
Uncond.	-0.004	-0.005	-0.003
	(0.004)	(0.004)	(0.004)
Cond. on Tier 1	-0.007	-0.008^{*}	-0.008^{*}
	(0.005)	(0.005)	(0.004)
Cond. on Tier 2	-0.007	-0.009^{**}	-0.008^{*}
	(0.005)	(0.005)	(0.004)
Cond. on chief complaint	-0.011^{**}	-0.012^{***}	-0.011^{**}
	(0.005)	(0.004)	(0.004)
Cond. on Tier 1 × age bin × gender	-0.010^{*}	-0.012^{**}	-0.011^{**}
	(0.006)	(0.005)	(0.005)
Cond. on Tier 2 × age bin × gender	-0.004	-0.006	-0.006
	(0.005)	(0.005)	(0.005)
Panel B: ICD-9 4-Digit Diagnosis			
Uncond.	-0.0003	-0.001	0.00002
	(0.003)	(0.003)	(0.003)
Cond. on Tier 1	-0.003	-0.004	-0.004
	(0.004)	(0.004)	(0.004)
Cond. on Tier 2	-0.003	-0.006	-0.004
	(0.004)	(0.004)	(0.004)
Cond. on chief complaint	-0.007^{*}	-0.008^{**}	-0.007^{*}
	(0.004)	(0.004)	(0.004)
Cond. on Tier 1 × age bin × gender	-0.008	-0.009^{*}	-0.008^{*}
	(0.005)	(0.005)	(0.004)
Cond. on Tier 2 × age bin × gender	-0.003	-0.005	-0.004
	(0.005)	(0.005)	(0.004)

Notes: The table presents OLS estimates of linear regressions of entropy reduction (see text for the formal definition) on the leave-out measure of cognitive load, which is standardized to have mean zero and standard deviation one. The reported coefficients correspond to those shown in columns (3) of Table 4. The unit of observation is the encounter, considering all orders placed within the first X minutes of the first physician–patient interaction, where X is indicated in the column headers.
All specifications control for prior cognitive load and the number of orders placed, and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).
Columns are grouped by the time window over which entropy reduction is computed (20, 30, or 60 minutes after the first physician–patient interaction). Panel A uses entropy reduction constructed at the diagnosis-ID level, whereas Panel B uses entropy reduction constructed at the ICD-9 4-digit level. Within each panel, rows correspond to different conditioning sets used in constructing the entropy measures: *Uncond.* (no conditioning), conditioning on tier 1 chief-complaint categories, tier 2 chief-complaint categories, chief complaint, tier 1 × age bin × gender, and tier 2 × age bin × gender. Age bins are defined as 18–34, 35–49, 50–69, and 70+ years.

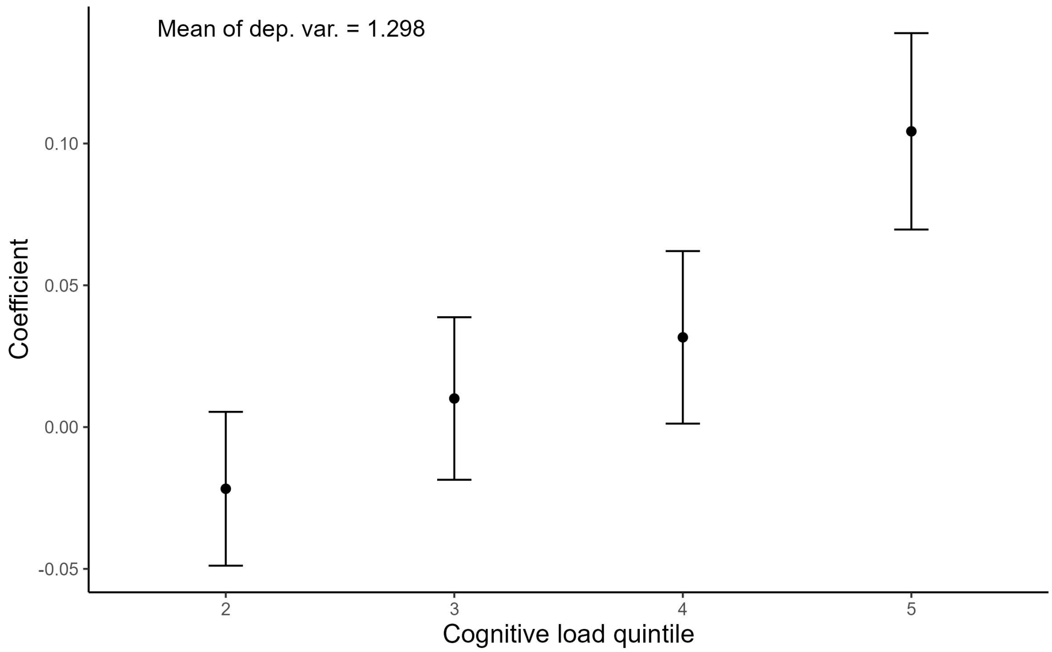
F.2 Alternative Functional Forms

F.2.1 Cognitive Load Quintiles

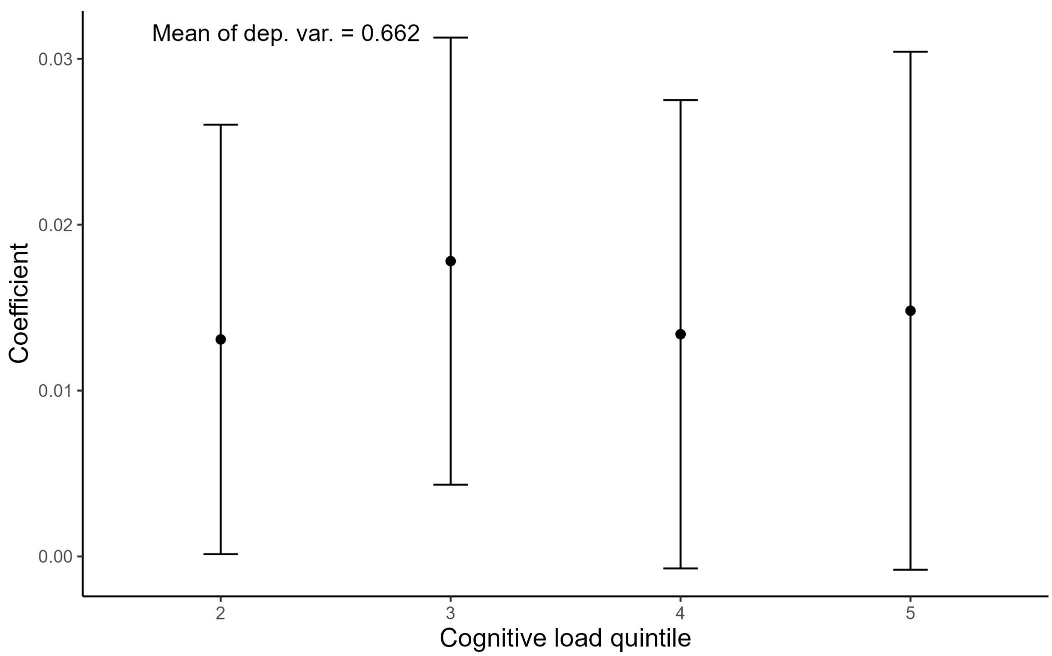
Table F.2: OLS Estimates of the Impact of Cognitive Load on Orders: Cognitive Load Quintiles

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Q5 cog. load	0.131***	0.097***	0.010	−0.006
	(0.014)	(0.015)	(0.006)	(0.007)
Q4 cog. load	0.039***	0.024*	0.005	−0.006
	(0.013)	(0.013)	(0.006)	(0.006)
Q2 cog. load	−0.044***	−0.030**	−0.014**	−0.005
	(0.013)	(0.013)	(0.006)	(0.006)
Q1 cog. load	−0.033**	−0.009	−0.018***	−0.017**
	(0.014)	(0.015)	(0.007)	(0.007)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,625	266,625	266,625	266,625
R^2	0.043	0.052	0.024	0.057
Adjusted R^2	0.040	0.047	0.021	0.052

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on quintiles of the leave-out measure of cognitive load. The unit of observation is the *encounter–event* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).



Panel A: Number of diagnostic orders



Panel B: Number of medication orders

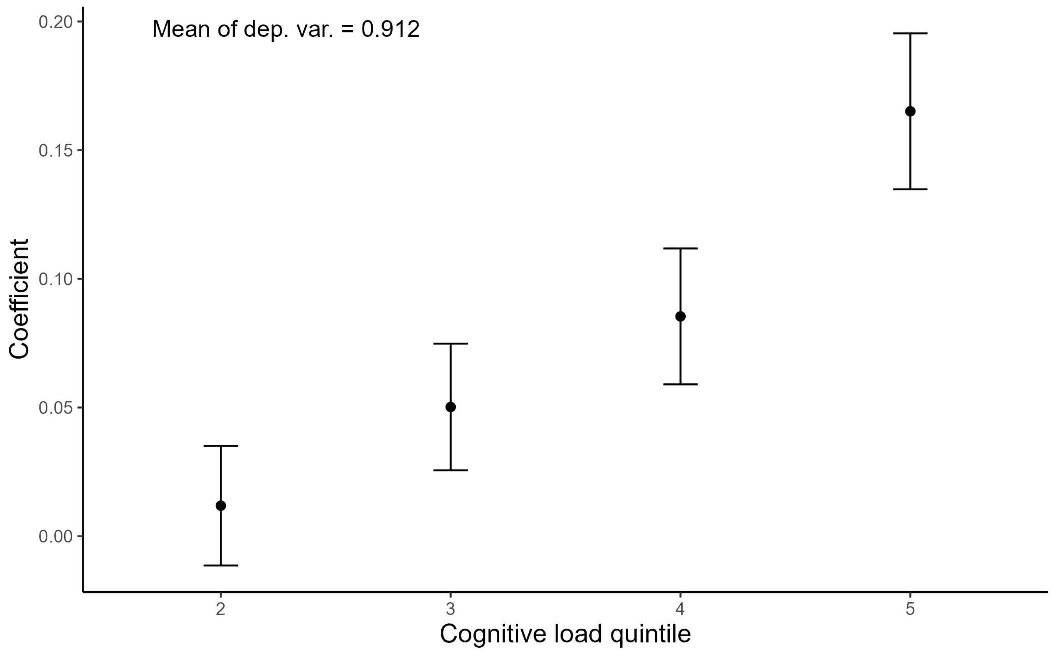
Figure F.1: Effect of Cognitive Load Quintiles on Number of Orders

Notes: The figure presents OLS estimates and 95% confidence intervals from linear regressions of the number of orders on indicators for quintiles of the leave-out cognitive load measure. Panel A shows coefficients for the number of diagnostic orders. Panel B shows coefficients for the number of medication orders. In both panels, coefficients for quintiles 2–5 are shown, with the lowest cognitive load quintile omitted. The unit of observation is the encounter–action level. All specifications control for prior cognitive load and include fixed effects for chief complaint, acuity code, hour of day, day of week, hours since shift start, hours until shift end, and physician. Standard errors are clustered at the encounter level.

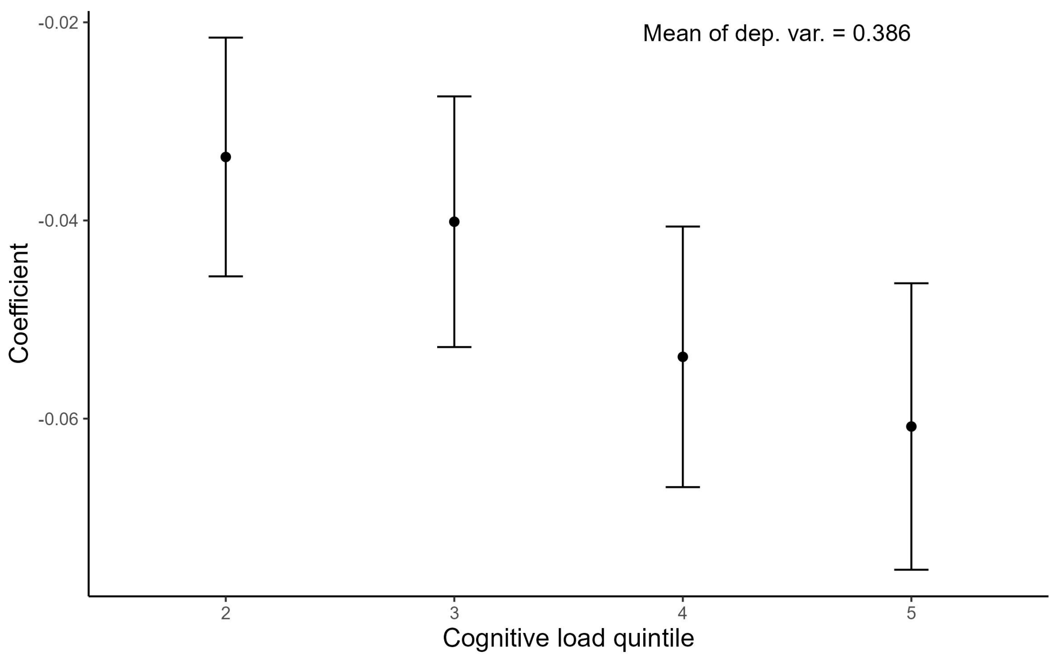
Table F.3: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking: Cognitive Load Quintiles

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent) editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Q5 cog. load	0.149***	0.117***	−0.017***	−0.020***	−0.325***	−0.113***
	(0.013)	(0.013)	(0.006)	(0.006)	(0.012)	(0.012)
Q4 cog. load	0.051***	0.038***	−0.012**	−0.014***	−0.135***	−0.052***
	(0.011)	(0.012)	(0.005)	(0.005)	(0.012)	(0.011)
Q2 cog. load	−0.047***	−0.036***	0.004	0.005	0.131***	0.065***
	(0.011)	(0.011)	(0.005)	(0.006)	(0.012)	(0.011)
Q1 cog. load	−0.076***	−0.047***	0.043***	0.038***	0.148***	0.103***
	(0.012)	(0.013)	(0.006)	(0.006)	(0.014)	(0.013)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.020	0.107

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on quintiles of the leave-out measure of cognitive load. The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).



Panel A: Number of top-25 diagnostic orders



Panel B: Number of non-top-25 diagnostic orders

Figure F.2: Effect of Cognitive Load Quintiles on Composition of Diagnostic Orders

Notes: The figure presents OLS estimates and 95% confidence intervals from linear regressions of the number of diagnostic orders on indicators for quintiles of the leave-out cognitive load measure. Panel A shows estimates for the number of diagnostic orders among the 25 most frequently placed diagnostic orders. Panel B shows estimates for the number of diagnostic orders outside the 25 most frequently placed diagnostic orders. In both panels, coefficients for quintiles 2–5 are shown, with the lowest cognitive load quintile omitted. The unit of observation is the encounter–action level. All specifications control for prior cognitive load and include fixed effects for chief complaint, acuity code, hour of day, day of week, hours since shift start, hours until shift end, and physician. Standard errors are clustered at the encounter level.

Table F.4: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision: Cognitive Load Quintiles

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Q5 cog. load	-0.017^{**}	-0.016^{*}	-0.016^{*}	-0.008
	(0.008)	(0.008)	(0.008)	(0.007)
Q4 cog. load	-0.007	-0.007	-0.004	-0.0003
	(0.008)	(0.008)	(0.008)	(0.007)
Q2 cog. load	0.001	0.0004	0.002	-0.003
	(0.008)	(0.008)	(0.008)	(0.007)
Q1 cog. load	0.043^{***}	0.012	0.013	0.002
	(0.008)	(0.009)	(0.009)	(0.008)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.62	0.62	0.62	0.62
N	56,398	56,398	56,398	56,398
R^2	0.345	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on quintiles of the leave-out measure of cognitive load. The unit of observation is the encounter, considering all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

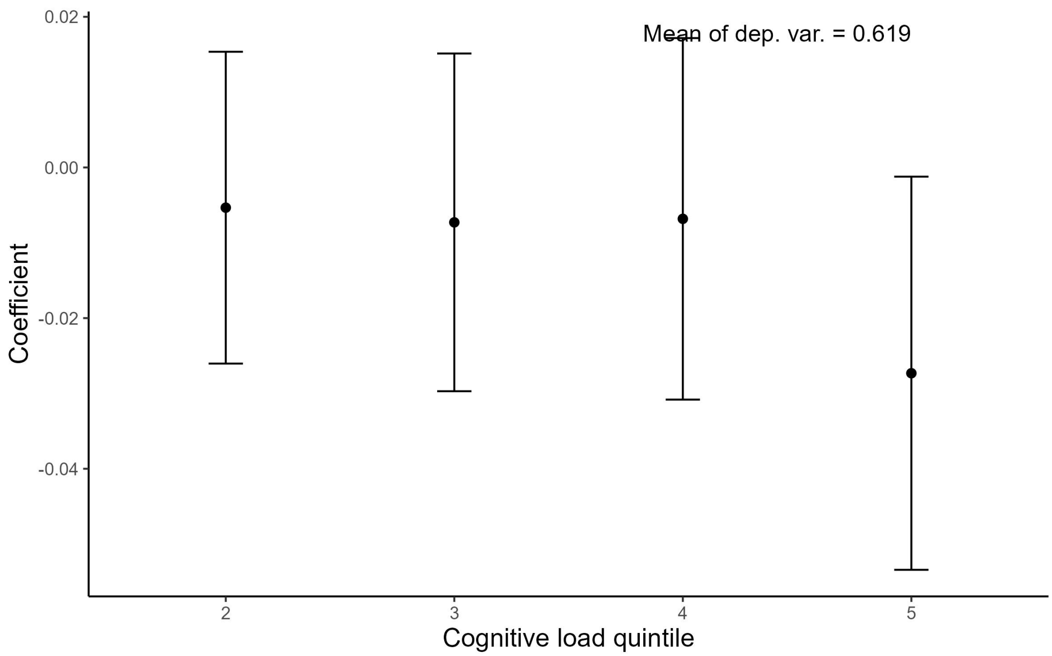


Figure F.3: Effect of Cognitive Load Quintiles on Entropy Reduction

Notes: The figure presents OLS estimates and 95% confidence intervals from linear regressions of entropy reduction (see text for definition) on indicators for quintiles of the leave-out cognitive load measure. Coefficients for quintiles 2-5 are shown, with the lowest cognitive load quintile omitted. The unit of observation is the encounter–action level. All specifications control for prior cognitive load and include fixed effects for chief complaint, acuity code, hour of day, day of week, hours since shift start, hours until shift end, and physician. Standard errors are clustered at the encounter level.

F.2.2 Logarithmic Cognitive Load

Table F.5: OLS Estimates of the Impact of Cognitive Load on Orders: Logarithmic Cognitive Load

	Number of diagnostic orders		Number medication orders	
	(1)	(2)	(3)	(4)
log(1 + Cognitive load)	0.036***	0.020***	0.009***	0.005*
	(0.005)	(0.006)	(0.003)	(0.003)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,625	266,625	266,625	266,625
R^2	0.042	0.052	0.024	0.057
Adjusted R^2	0.039	0.047	0.021	0.052

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the natural logarithm of the leave-out measure of cognitive load. The unit of observation is the *encounter–event* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.6: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking: Logarithmic Cognitive Load

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
log(1 + Cognitive load)	0.059***	0.041***	−0.023***	−0.021***	−0.133***	−0.065***
	(0.004)	(0.005)	(0.002)	(0.002)	(0.005)	(0.005)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.045	0.054	0.022	0.032	0.020	0.112
Adjusted R^2	0.042	0.049	0.019	0.027	0.017	0.107

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the natural logarithm of the leave-out measure of cognitive load. The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.7: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision: Logarithmic Cognitive Load

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
log(1 + Cognitive load)	-0.018^{***}	-0.006^{*}	-0.006^{*}	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.62	0.62	0.62	0.62
N	56,398	56,398	56,398	56,398
R^2	0.345	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the natural logarithm of the leave-out measure of cognitive load. The unit of observation is the encounter, considering all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.2.3 Logarithmic Number of Orders

Table F.8: OLS Estimates of the Impact of Cognitive Load on Logarithmic Orders

	log(1 + Number of diagnostic orders)		log(1 + Number of medication orders)	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.017***	0.011***	0.011***	0.005***
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,609	266,609	266,609	266,609
R^2	0.054	0.061	0.033	0.049
Adjusted R^2	0.051	0.056	0.030	0.044

Notes: The table presents OLS estimates of linear regressions of the logarithmic number of diagnostic and medication orders on the leave-out measure of cognitive load. The unit of observation is the *encounter–event* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.9: OLS Estimates of the Impact of Cognitive Load on Logarithmic Frequent Orders

	log(1 + Number of top25 DX orders)		log(1 + Number of non-top25 DX orders)	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.024***	0.019***	−0.001*	−0.002**
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.91	0.91	0.14	0.14
N	266,625	266,625	266,625	266,625
R^2	0.054	0.061	0.015	0.026
Adjusted R^2	0.052	0.056	0.013	0.021

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.3 First-Order Batch

F.3.1 Orders Placed Within 20 minutes

Table F.10: OLS Estimates of the Impact of Cognitive Load on Orders Placed Within First 20 Minutes

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.200***	0.024	0.026***	−0.015**
	(0.020)	(0.021)	(0.006)	(0.007)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.07	4.07	0.86	0.86
N	76,001	76,001	76,001	76,001
R^2	0.249	0.274	0.105	0.136
Adjusted R^2	0.241	0.260	0.096	0.119

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders placed within the first 20 minutes of the first physician-patient interaction on the leave-out measure of cognitive load. The unit of observation is at the *encounter* level. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.11: OLS Estimates of the Impact of Cognitive Load on Frequent Orders Placed Within First 20 Minutes

	Number of top25 DX orders		Number of non-top25 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.211***	0.044**	−0.011	−0.020**
	(0.016)	(0.017)	(0.009)	(0.009)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	3.15	3.15	0.92	0.92
N	76,001	76,001	76,001	76,001
R^2	0.265	0.294	0.068	0.088
Adjusted R^2	0.258	0.281	0.058	0.071

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed orders, placed within the first 20 minutes of the first physician-patient interaction, on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is at the *encounter* level. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.3.2 Orders Placed Within 30 minutes

Table F.12: OLS Estimates of the Impact of Cognitive Load on Orders Placed Within First 30 Minutes

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.217***	0.033	0.033***	−0.015**
	(0.021)	(0.022)	(0.007)	(0.008)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.36	4.36	0.99	0.99
N	76,001	76,001	76,001	76,001
R^2	0.263	0.288	0.112	0.142
Adjusted R^2	0.255	0.275	0.103	0.126

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders placed within the first 30 minutes of the first physician-patient interaction on the leave-out measure of cognitive load. The unit of observation is at the *encounter* level. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.13: OLS Estimates of the Impact of Cognitive Load on Frequent Orders Placed Within First 30 Minutes

	Number of top25 DX orders		Number of non-top25 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.222***	0.050***	−0.006	−0.016*
	(0.017)	(0.018)	(0.009)	(0.010)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	3.35	3.35	1.01	1.01
N	76,001	76,001	76,001	76,001
R^2	0.280	0.310	0.074	0.095
Adjusted R^2	0.273	0.297	0.065	0.078

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed orders, placed within the first 30 minutes of the first physician-patient interaction, on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is at the *encounter* level. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.3.3 Orders Placed Within 60 minutes

Table F.14: OLS Estimates of the Impact of Cognitive Load on Orders Placed Within First 60 Minutes

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.231***	0.038	0.045***	−0.014
	(0.022)	(0.024)	(0.008)	(0.009)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.99	4.99	1.33	1.33
N	76,001	76,001	76,001	76,001
R^2	0.284	0.312	0.125	0.153
Adjusted R^2	0.276	0.300	0.115	0.138

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders placed within the first 60 minutes of the first physician-patient interaction on the leave-out measure of cognitive load. The unit of observation is at the *encounter* level. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.15: OLS Estimates of the Impact of Cognitive Load on Frequent Orders Placed Within First 60 Minutes

	Number of top25 DX orders		Number of non-top25 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.233***	0.054***	−0.002	−0.016
	(0.017)	(0.019)	(0.010)	(0.011)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	3.77	3.77	1.22	1.22
N	76,001	76,001	76,001	76,001
R^2	0.303	0.335	0.087	0.109
Adjusted R^2	0.296	0.323	0.077	0.092

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders and the number of orders outside the 25 most frequently placed orders, placed within the first 60 minutes of the first physician-patient interaction, on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is at the *encounter* level. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.4 Alternative Cognitive Load Windows

F.4.1 60 Minutes Window

Table F.16: OLS Estimates of the Impact of Cognitive Load Based on 60-Minutes Window on Orders

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.) (60 min.)	0.052***	0.036***	0.004*	−0.002
	(0.006)	(0.006)	(0.002)	(0.003)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,625	266,625	266,625	266,625
R^2	0.042	0.052	0.024	0.057
Adjusted R^2	0.040	0.047	0.021	0.052

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 60 minutes (instead of 90 minutes, as in the standard specification). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.17: OLS Estimates of the Impact of Cognitive Load Based on 60-Minutes Window on Frequent Orders and Note Taking

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.) (60 min.)	0.073^{***}	0.056^{***}	-0.021^{***}	-0.021^{***}	-0.153^{***}	-0.067^{***}
	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)	(0.004)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.019	0.107

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 60 minutes (instead of 90 minutes as in the standard specification). The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.18: OLS Estimates of the Impact of Cognitive Load Based on 60-Minutes Window on Diagnostic Precision

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.) (60 min.)	-0.012^{***}	-0.006^{*}	-0.006^{**}	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.48	0.48	0.48	0.48
N	48,211	48,211	48,211	48,211
R^2	0.344	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 60 minutes (instead of 90 minutes as in the standard specification). The unit of observation are all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.4.2 120 Minutes Window

Table F.19: OLS Estimates of the Impact of Cognitive Load Based on 120-Minutes Window on Orders

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.) (120 min.)	0.077***	0.052***	0.015***	0.006**
	(0.006)	(0.006)	(0.003)	(0.003)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,625	266,625	266,625	266,625
R^2	0.043	0.052	0.024	0.057
Adjusted R^2	0.040	0.047	0.021	0.052

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 120 minutes (instead of 90 minutes as in the standard specification). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.20: OLS Estimates of the Impact of Cognitive Load Based on 120-Minutes Window on Frequent Orders and Note Taking

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.) (120 min.)	0.096^{***}	0.070^{***}	-0.019^{***}	-0.018^{***}	-0.194^{***}	-0.078^{***}
	(0.005)	(0.006)	(0.002)	(0.003)	(0.005)	(0.005)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.046	0.054	0.022	0.032	0.023	0.112
Adjusted R^2	0.043	0.049	0.019	0.027	0.019	0.107

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 120 minutes (instead of 90 minutes as in the standard specification). The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.21: OLS Estimates of the Impact of Cognitive Load Based on 120-Minutes Window on Diagnostic Precision

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.) (120 min.)	-0.019^{***}	-0.008^{**}	-0.008^{**}	-0.002
	(0.003)	(0.004)	(0.004)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.62	0.62	0.62	0.62
N	56,398	56,398	56,398	56,398
R^2	0.344	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The relevant window for the computation of cognitive load is now 120 minutes (instead of 90 minutes as in the standard specification). The unit of observation are all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.5 Cognitive Load Based on Past Orders Only

Table F.22: OLS Estimates of the Impact of Cognitive Load (Based on Past Orders Only) on Orders

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.073***	0.048***	0.005**	0.003
	(0.006)	(0.006)	(0.003)	(0.003)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.30	1.30	0.66	0.66
N	266,625	266,625	266,625	266,625
R^2	0.042	0.052	0.024	0.057
Adjusted R^2	0.040	0.047	0.021	0.052

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). Only orders (instead of orders *and* notes as in the standard specification) within the last 90 minutes enter the measure cognitive load. The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.23: OLS Estimates of the Impact of Cognitive Load (Based on Past Orders Only) on Frequent Orders and Note Taking

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.091^{***}	0.065^{***}	-0.018^{***}	-0.017^{***}	-0.138^{***}	-0.047^{***}
	(0.005)	(0.005)	(0.002)	(0.002)	(0.004)	(0.004)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	0.91	0.91	0.39	0.39	446.50	446.50
N	266,625	266,625	266,625	266,625	230,682	230,682
R^2	0.045	0.054	0.022	0.032	0.023	0.111
Adjusted R^2	0.043	0.049	0.019	0.027	0.020	0.107

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). Only orders (instead of orders *and* notes as in the standard specification) within the last 90 minutes enter the measure cognitive load. The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.24: OLS Estimates of the Impact of Cognitive Load (Based on Past Orders Only) on Diagnostic Precision

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.018^{***}	-0.007^{**}	-0.007^{**}	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Number of DX orders			-0.021^{***}	-0.020^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.381^{***}
				(0.026)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.62	0.62	0.62	0.62
N	56,398	56,398	56,398	56,398
R^2	0.344	0.369	0.383	0.510
Adjusted R^2	0.335	0.352	0.367	0.497

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). Only orders (instead of orders *and* notes as in the standard specification) within the last 90 minutes enter the measure cognitive load. The unit of observation are all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.6 Abnormal orders

Table F.25: OLS Estimates of the Impact of Cognitive Load on Share of Abnormal Orders

	Share of orders with abnormal test result	
	(1)	(2)
Cognitive load (std.)	0.0003	0.001
	(0.001)	(0.001)
Prior cognitive load - Control	X	X
Hour of day - FE	X	X
Day of week - FE	X	X
Hours since shift start - FE	X	X
Hours until shift end - FE	X	X
Acuity code - FE	X	X
Chief complaint - FE	X	X
Physician - FE		X
Mean of DV	0.06	0.06
N	266,625	266,625
R^2	0.012	0.017
Adjusted R^2	0.009	0.012

Notes: The table presents OLS estimates of linear regressions of the share of diagnostic orders that yield an abnormal result on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Column (2) additionally includes physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.7 Varying Threshold for Common Orders

F.7.1 10 Most Frequent Orders

Table F.26: OLS Estimates of the Impact of Cognitive Load on Top10 Frequent Orders

	Number of top10 DX orders		Number of non-top10 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.056***	0.046***	0.015***	0.003
	(0.003)	(0.003)	(0.004)	(0.004)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.57	0.57	0.72	0.72
N	266,625	266,625	266,625	266,625
R^2	0.045	0.053	0.028	0.039
Adjusted R^2	0.042	0.048	0.025	0.034

Notes: The table presents OLS estimates of linear regressions of the number of the 10 most frequently placed orders and the number of orders outside the 10 most frequently placed orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.7.2 50 Most Frequent Orders

Table F.27: OLS Estimates of the Impact of Cognitive Load on Top50 Frequent Orders

	Number of top50 DX orders		Number of non-top50 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.083***	0.063***	−0.012***	−0.013***
	(0.005)	(0.006)	(0.002)	(0.002)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.05	1.05	0.24	0.24
N	266,625	266,625	266,625	266,625
R^2	0.045	0.054	0.027	0.036
Adjusted R^2	0.043	0.049	0.025	0.031

Notes: The table presents OLS estimates of linear regressions of the number of the 50 most frequently placed orders and the number of orders outside the 50 most frequently placed orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.7.3 100 Most Frequent Orders

Table F.28: OLS Estimates of the Impact of Cognitive Load on Top100 Frequent Orders

	Number of top100 DX orders		Number of non-top100 DX orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.080***	0.060***	−0.010***	−0.010***
	(0.005)	(0.006)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.16	1.16	0.14	0.14
N	266,625	266,625	266,625	266,625
R^2	0.047	0.056	0.034	0.040
Adjusted R^2	0.044	0.051	0.031	0.035

Notes: The table presents OLS estimates of linear regressions of the number of the 100 most frequently placed orders and the number of orders outside the 100 most frequently placed orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1). The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

F.8 Individual Chief Complaints

F.8.1 Abdominal Pain

Table F.29: OLS Estimates of the Impact of Cognitive Load on Orders for Encounters with Abdominal Pain

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.066***	0.053***	0.027***	0.007
	(0.018)	(0.020)	(0.007)	(0.008)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.52	1.52	0.72	0.72
N	33,185	33,185	33,185	33,185
R^2	0.035	0.056	0.007	0.036
Adjusted R^2	0.033	0.042	0.004	0.022

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders and of the share of diagnostic orders that yield an abnormal result on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Abdominal Pain." The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.30: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking for Encounters with Abdominal Pain

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.125^{***}	0.099^{***}	-0.059^{***}	-0.046^{***}	-0.200^{***}	-0.090^{***}
	(0.017)	(0.018)	(0.007)	(0.008)	(0.014)	(0.014)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.14	1.14	0.39	0.39	435.53	435.53
N	33,185	33,185	33,185	33,185	25,715	25,715
R^2	0.037	0.056	0.010	0.036	0.024	0.124
Adjusted R^2	0.034	0.043	0.008	0.022	0.021	0.109

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Abdominal Pain." The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

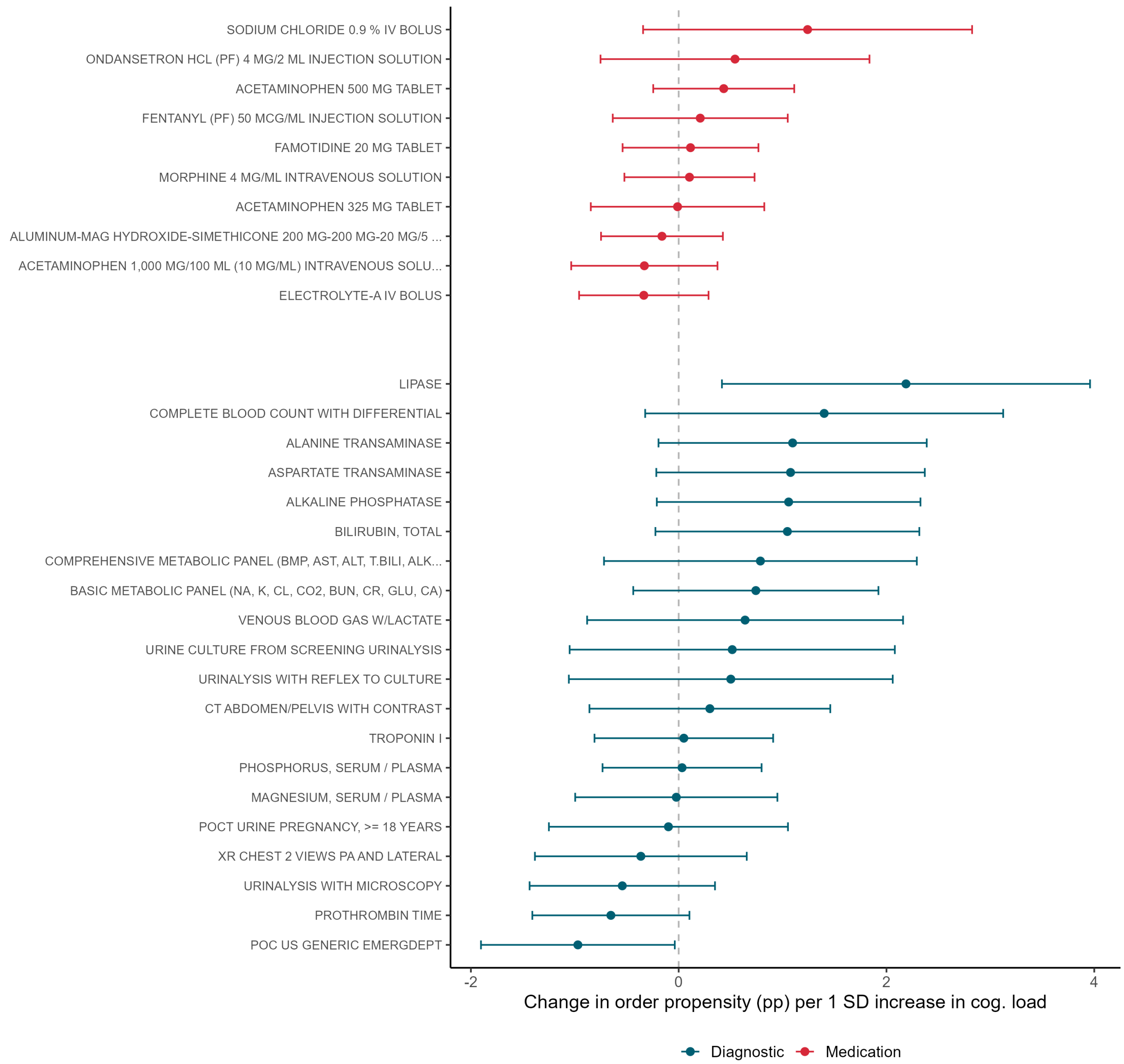


Figure F.4: Effect of Cognitive Load on Top-30 Individual Orders for Abdominal Pain Encounters

Notes: The figure presents OLS estimates of linear probability models of indicators for whether each of the 30 most frequently placed individual orders is used on the leave-out cognitive load measure (standardized to mean zero and standard deviation of one), for the subset of encounters with chief complaint "Abdominal Pain." The unit of observation is the encounter-order level. All specifications control for prior cognitive load and include fixed effects for physician, hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the physician level and shown as 95% confidence intervals.

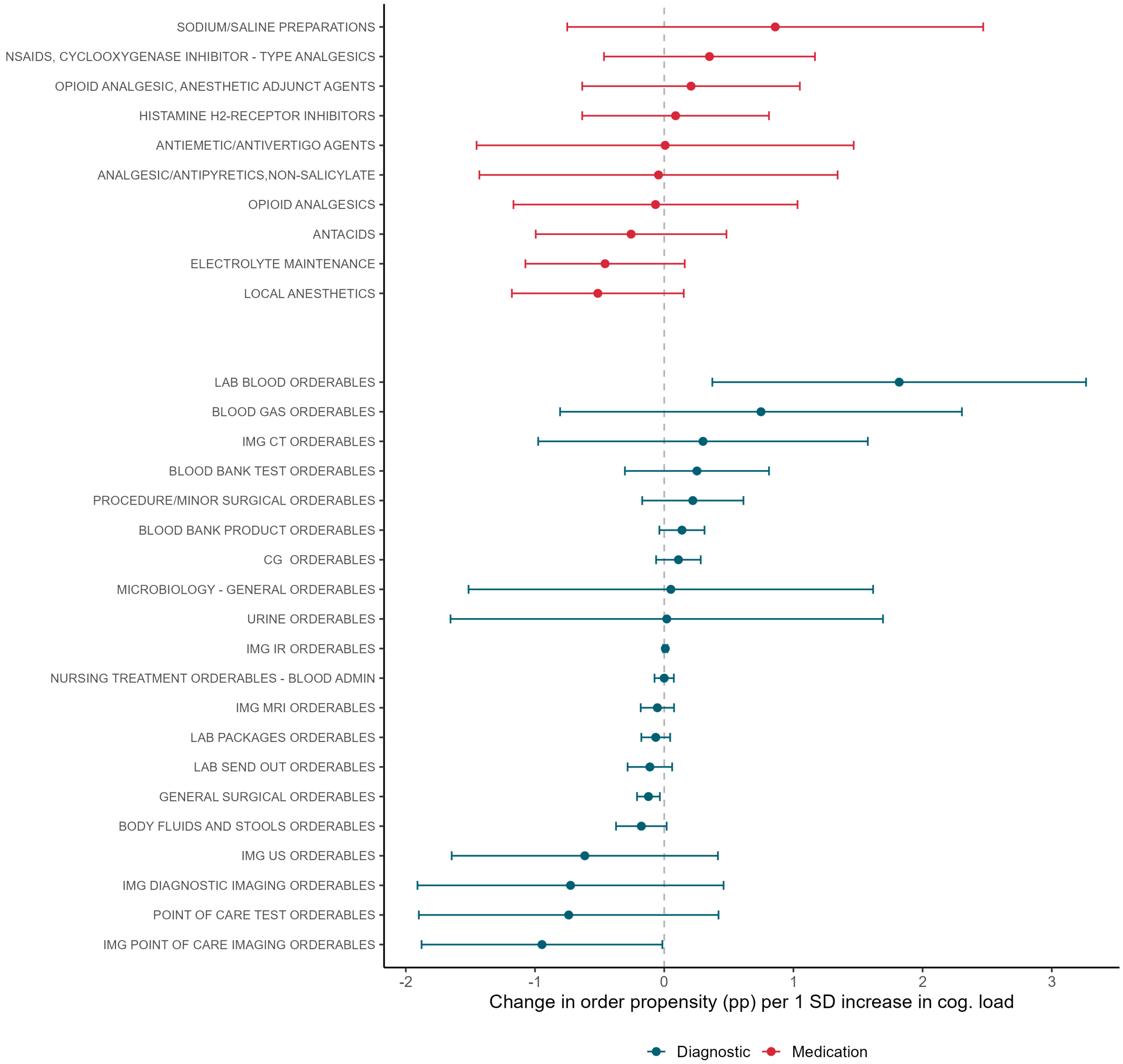


Figure F.5: Effect of Cognitive Load on Top-30 Order Subcategories for Abdominal Pain Encounters

Notes: The figure presents OLS estimates of linear probability models of indicators for whether each of the 30 most frequently used order subcategories is placed on the leave-out cognitive load measure (standardized to mean zero and standard deviation of one), for the subset of encounters with chief complaint "Abdominal Pain." The unit of observation is the encounter-order level. All specifications control for prior cognitive load and include fixed effects for physician, hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the physician level and shown as 95% confidence intervals.

Table F.31: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision for Encounters with Abdominal Pain

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.021^{***}	-0.008	-0.009	-0.003
	(0.005)	(0.006)	(0.006)	(0.005)
Number of DX orders			-0.010^{***}	-0.009^{***}
			(0.001)	(0.001)
Share of top50 DX orders				-1.267^{***}
				(0.067)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.24	0.24	0.24	0.24
N	6,837	6,837	6,837	6,837
R^2	0.025	0.137	0.151	0.387
Adjusted R^2	0.013	0.076	0.091	0.343

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Abdominal Pain". The unit of observation are all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

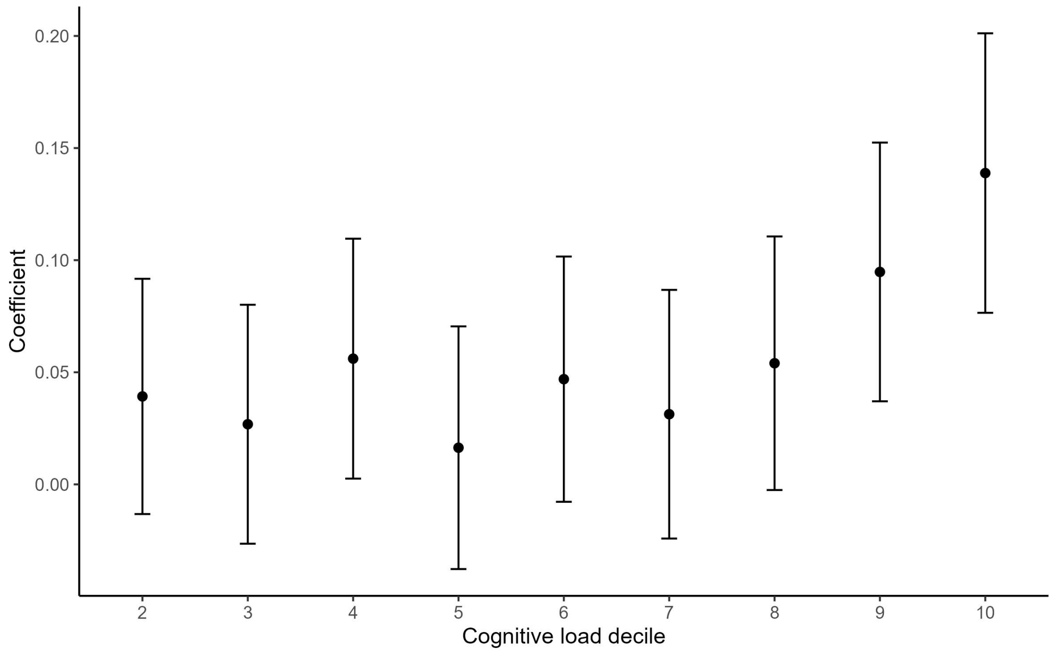


Figure F.6: Effect of Cognitive Load Deciles on Inpatient Admission (Abdominal Pain)

Notes: The figure plots OLS estimates and 95% confidence intervals from linear probability models of a dummy variable for whether the patient was admitted to the hospital during the encounter on indicators for deciles of the maximum leave-out cognitive load measure during the encounter, for the subset of encounters with chief complaint "Abdominal Pain." Coefficients for deciles 2–10 are shown, with the lowest cognitive load decile omitted. The unit of observation is the encounter. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, acuity code, and physician. Standard errors are heteroskedasticity-robust.

F.8.2 Chest Pain

Table F.32: OLS Estimates of the Impact of Cognitive Load on Orders for Encounters with Chest Pain

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Cognitive load (std.)	0.105***	0.103***	−0.028**	−0.020*
	(0.022)	(0.024)	(0.011)	(0.011)
Prior cognitive load - Control	X	X	X	X
Hour of day - FE	X	X	X	X
Day of week - FE	X	X	X	X
Hours since shift start - FE	X	X	X	X
Hours until shift end - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	1.44	1.44	0.59	0.59
N	17,295	17,295	17,295	17,295
R^2	0.037	0.070	0.010	0.096
Adjusted R^2	0.032	0.045	0.005	0.071

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Chest Pain." The unit of observation is the *encounter–action* level, where an *action* refers to a placed batch of orders. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table F.33: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking for Encounters with Chest Pain

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Cognitive load (std.)	0.126***	0.116***	−0.020**	−0.013	−0.181***	−0.080***
	(0.020)	(0.022)	(0.008)	(0.008)	(0.017)	(0.017)
Prior cognitive load - Control	X	X	X	X	X	X
Randomization block - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	1.11	1.11	0.32	0.32	464.20	464.20
N	17,295	17,295	17,295	17,295	15,542	15,542
R^2	0.041	0.070	0.006	0.046	0.026	0.137
Adjusted R^2	0.036	0.046	0.001	0.020	0.020	0.114

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Chest Pain." The unit of observation is the *encounter–action* level, where an *action* either refers to a placed batch of orders (columns (1)-(4) or a note-taking instance (columns (5)-(6)). All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

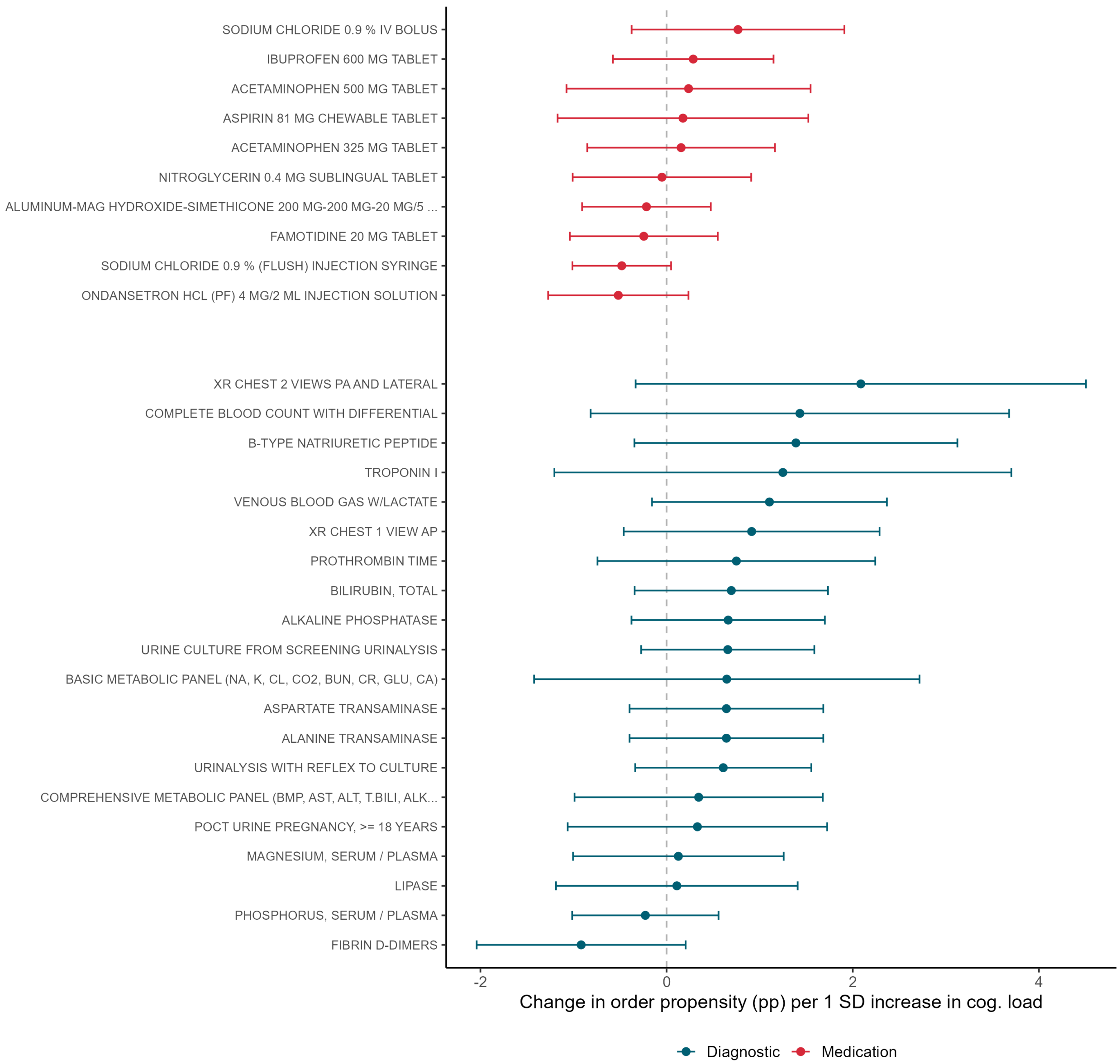


Figure F.7: Effect of Cognitive Load on Top-30 Individual Orders for Chest Pain Encounters

Notes: The figure presents OLS estimates of linear probability models of indicators for whether each of the 30 most frequently placed individual orders is used on the leave-out cognitive load measure (standardized to mean zero and standard deviation of one), for the subset of encounters with chief complaint "Chest Pain." The unit of observation is the encounter-order level. All specifications control for prior cognitive load and include fixed effects for physician, hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the physician level and shown as 95% confidence intervals.

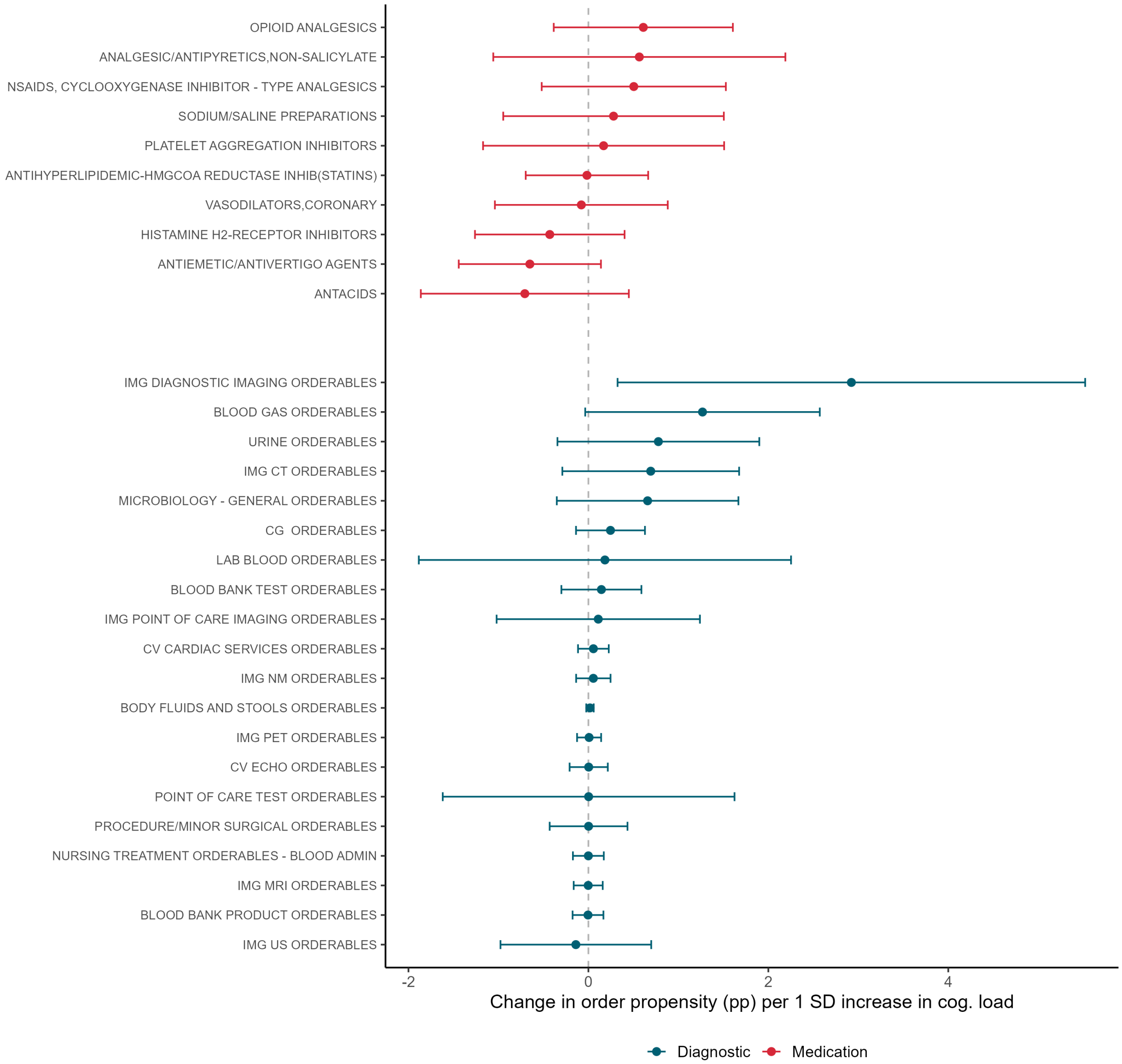


Figure F.8: Effect of Cognitive Load on Top-30 Order Subcategories for Chest Pain Encounters

Notes: The figure presents OLS estimates of linear probability models of indicators for whether each of the 30 most frequently used order subcategories is placed on the leave-out cognitive load measure (standardized to mean zero and standard deviation of one), for the subset of encounters with chief complaint "Chest Pain." The unit of observation is the encounter-order level. All specifications control for prior cognitive load and include fixed effects for physician, hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the physician level and shown as 95% confidence intervals.

Table F.34: OLS Estimates of the Impact of Cognitive Load on Diagnostic Precision for Encounters with Chest Pain

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Cognitive load (std.)	-0.022^{**}	0.001	0.005	0.004
	(0.009)	(0.010)	(0.010)	(0.009)
Number of DX orders			-0.026^{***}	-0.027^{***}
			(0.002)	(0.002)
Share of top50 DX orders				-1.913^{***}
				(0.173)
Prior cognitive load - Control	X	X	X	X
Randomization block - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.16	0.16	0.16	0.16
N	3,452	3,452	3,452	3,452
R^2	0.047	0.192	0.239	0.431
Adjusted R^2	0.023	0.082	0.134	0.352

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the leave-out measure of cognitive load (standardized to mean 0 and SD 1), for the subset of encounters with chief complaint "Chest Pain". The unit of observation are all orders placed within 20 minutes of the first physician-patient interaction. All specifications control for prior cognitive load and include fixed effects for hour of day, day of week, hours since shift start, hours until shift end, chief complaint, and acuity code. Standard errors are clustered at the encounter level and shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

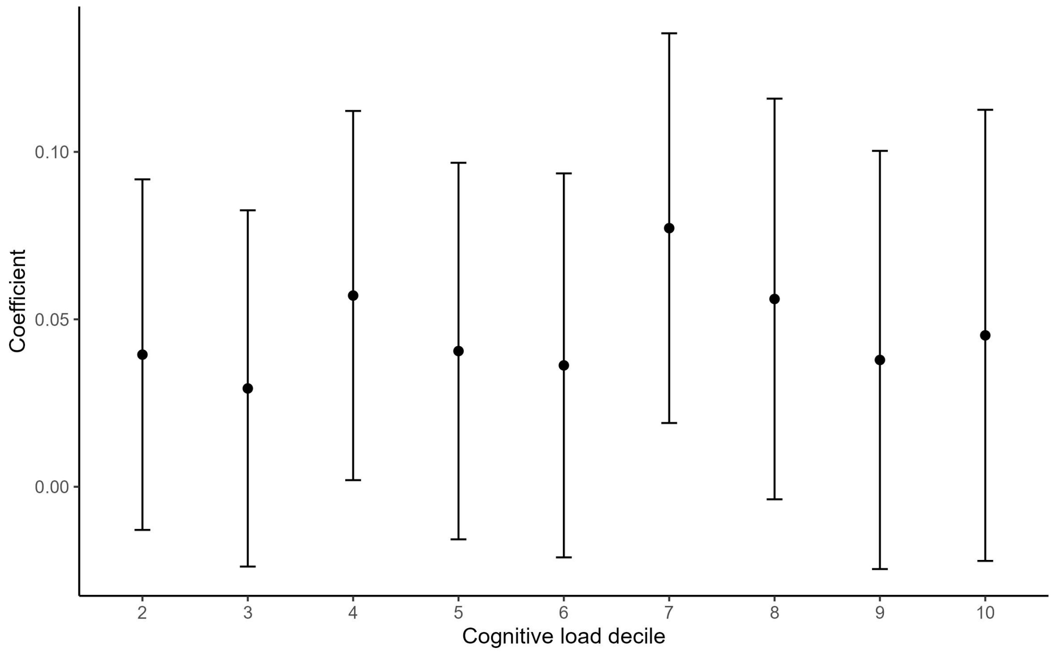


Figure F.9: Effect of Cognitive Load Deciles on Inpatient Admission (Chest Pain)

Notes: The figure plots OLS estimates and 95% confidence intervals from linear probability models of a dummy variable for whether the patient was admitted to the hospital during the encounter on indicators for deciles of the maximum leave-out cognitive load measure during the encounter, for the subset of encounters with chief complaint “Chest Pain.” Coefficients for deciles 2–10 are shown, with the lowest cognitive load decile omitted. The unit of observation is the encounter. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, acuity code, and physician. Standard errors are heteroskedasticity-robust.

G Alternative Regression Specifications: Encounter Level

G.1 Max Cognitive Load

Table G.1: OLS Estimates of the Impact of Cognitive Load on Orders at the Encounter-Level

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Max cognitive load (std.)	0.701***	0.604***	0.345***	0.321***
	(0.022)	(0.024)	(0.012)	(0.013)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.55	4.55	2.32	2.32
N	76,001	76,001	76,001	76,001
R^2	0.201	0.236	0.142	0.211
Adjusted R^2	0.193	0.222	0.134	0.197

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.2: OLS Estimates of the Impact of Cognitive Load on Frequent Orders and Note Taking at the Encounter-Level

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Max cognitive load (std.)	0.531***	0.423***	0.171***	0.181***	0.264***	0.135***
	(0.017)	(0.019)	(0.010)	(0.011)	(0.011)	(0.010)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	3.20	3.20	1.35	1.35	1355.24	1355.24
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.201	0.236	0.102	0.134	0.061	0.361
Adjusted R^2	0.193	0.222	0.093	0.119	0.052	0.350

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.3: OLS Estimates of the Impact of Cognitive Load on Entropy Reduction at the Encounter-Level

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Max cognitive load (std.)	-0.035^{***}	-0.026^{***}	-0.019^{***}	-0.017^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Number of diagnostic orders			-0.014^{***}	-0.007^{***}
			(0.0005)	(0.0005)
Share of top50 DX orders				-0.344^{***}
				(0.007)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.71	0.71	0.71	0.71
N	57,048	57,048	57,048	57,048
R^2	0.391	0.411	0.420	0.447
Adjusted R^2	0.384	0.398	0.407	0.434

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (3), and (4) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.4: OLS Estimates of the Impact of Deciles of Cognitive Load on Number of DX Orders and Inpatient Admission at the Encounter

	Number of diagnostic orders		Inpatient admission	
	(1)	(2)	(3)	(4)
Max cog. load decile 2	0.392***	0.198***	−0.006	0.008
	(0.062)	(0.070)	(0.005)	(0.006)
Max cog. load decile 3	0.841***	0.484***	0.001	0.007
	(0.065)	(0.073)	(0.006)	(0.007)
Max cog. load decile 4	1.116***	0.756***	0.012**	0.022***
	(0.067)	(0.075)	(0.006)	(0.007)
Max cog. load decile 5	1.363***	0.933***	0.014**	0.016**
	(0.070)	(0.077)	(0.006)	(0.007)
Max cog. load decile 6	1.566***	1.155***	0.026***	0.030***
	(0.072)	(0.078)	(0.006)	(0.007)
Max cog. load decile 7	1.762***	1.312***	0.038***	0.037***
	(0.074)	(0.080)	(0.007)	(0.007)
Max cog. load decile 8	1.961***	1.513***	0.044***	0.044***
	(0.076)	(0.082)	(0.007)	(0.007)
Max cog. load decile 9	2.056***	1.615***	0.053***	0.052***
	(0.079)	(0.085)	(0.007)	(0.008)
Max cog. load decile 10	2.341***	1.930***	0.067***	0.067***
	(0.085)	(0.092)	(0.008)	(0.008)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.55	4.55	0.24	0.24
N	76,001	76,001	76,001	76,001
R^2	0.202	0.236	0.221	0.242
Adjusted R^2	0.195	0.222	0.213	0.228

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic orders and a dummy variable whether the patient was admitted to the hospital during the encounter on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

G.2 Mean Cognitive Load

Table G.5: OLS Estimates of the Impact of Mean Cognitive Load on Orders at the Encounter-Level

	Number of diagnostic orders		Number of medication orders	
	(1)	(2)	(3)	(4)
Mean cognitive load (std.)	0.285***	−0.005	0.122***	−0.023*
	(0.020)	(0.023)	(0.011)	(0.012)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.55	4.55	2.32	2.32
N	76,001	76,001	76,001	76,001
R^2	0.197	0.243	0.141	0.219
Adjusted R^2	0.190	0.229	0.133	0.205

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic and medication orders on the mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2) and (4) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.6: OLS Estimates of the Impact of Mean Cognitive Load on Frequent Orders and Note Taking at the Encounter-Level

	Number of top25 DX orders		Number of non-top25 DX orders		log(Time spent editing notes)	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean cognitive load (std.)	0.275^{***}	0.037^{**}	0.011	-0.042^{***}	0.166^{***}	0.002
	(0.016)	(0.018)	(0.009)	(0.011)	(0.011)	(0.009)
Prior cognitive load - Control	X	X	X	X	X	X
Max hours since shift start - FE	X	X	X	X	X	X
Acuity code - FE	X	X	X	X	X	X
Chief complaint - FE	X	X	X	X	X	X
Physician - FE		X		X		X
Mean of DV	3.20	3.20	1.35	1.35	1355.24	1355.24
N	76,001	76,001	76,001	76,001	76,001	76,001
R^2	0.193	0.239	0.108	0.143	0.050	0.360
Adjusted R^2	0.185	0.225	0.100	0.127	0.041	0.349

Notes: The table presents OLS estimates of linear regressions of the number of the 25 most frequently placed orders, the number of orders outside the 25 most frequently placed orders, and the log of time spent editing the patient's note (in seconds) on the mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.7: OLS Estimates of the Impact of Mean Cognitive Load on Entropy Reduction at the Encounter-Level

	Entropy reduction (cond. on cc; DX orders)			
	(1)	(2)	(3)	(4)
Mean cognitive load (std.)	-0.026^{***}	-0.013^{***}	-0.013^{***}	-0.011^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Number of diagnostic orders			-0.014^{***}	-0.008^{***}
			(0.0005)	(0.0005)
Share of top50 DX orders				-0.345^{***}
				(0.007)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X	X	X
Mean of DV	0.71	0.71	0.71	0.71
N	57,048	57,048	57,048	57,048
R^2	0.389	0.410	0.420	0.446
Adjusted R^2	0.382	0.397	0.407	0.434

Notes: The table presents OLS estimates of linear regressions of the entropy reduction (see text for definition) on the mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (3), and (4) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.8: OLS Estimates of the Impact of Mean Cognitive Load on Inpatient Admission and Readmission at the Encounter-Level

	Inpatient admission		Readmission within 30 days	
	(1)	(2)	(3)	(4)
Mean cognitive load (std.)	0.001	-0.010^{***}	-0.002^{*}	-0.003^{***}
	(0.002)	(0.002)	(0.001)	(0.001)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	0.24	0.24	0.04	0.04
N	76,001	76,001	76,001	76,001
R^2	0.220	0.243	0.048	0.062
Adjusted R^2	0.212	0.229	0.039	0.045

Notes: The table presents OLS estimates of linear regressions of dummy variables whether the patient was admitted to the hospital during the encounter or re-admitted within 30 days on the mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

Table G.9: OLS Estimates of the Impact of Deciles of Mean Cognitive Load on Number of DX Orders and Inpatient Admission at the Encounter

	Number of diagnostic orders		Inpatient admission	
	(1)	(2)	(3)	(4)
Mean cog. load decile 2	1.358***	1.020***	0.024***	0.034***
	(0.069)	(0.070)	(0.006)	(0.006)
Mean cog. load decile 3	0.905***	0.357***	0.001	0.005
	(0.065)	(0.072)	(0.006)	(0.007)
Mean cog. load decile 4	1.657***	0.960***	0.025***	0.022***
	(0.069)	(0.074)	(0.006)	(0.007)
Mean cog. load decile 5	1.506***	0.729***	0.021***	0.016**
	(0.069)	(0.075)	(0.006)	(0.007)
Mean cog. load decile 6	1.518***	0.678***	0.014**	0.003
	(0.070)	(0.077)	(0.006)	(0.007)
Mean cog. load decile 7	1.637***	0.724***	0.032***	0.018**
	(0.071)	(0.078)	(0.006)	(0.007)
Mean cog. load decile 8	1.557***	0.580***	0.017**	−0.0005
	(0.072)	(0.080)	(0.007)	(0.007)
Mean cog. load decile 9	1.603***	0.579***	0.021***	−0.0004
	(0.074)	(0.083)	(0.007)	(0.008)
Mean cog. load decile 10	1.519***	0.440***	0.008	−0.017**
	(0.078)	(0.089)	(0.007)	(0.008)
Prior cognitive load - Control	X	X	X	X
Max hours since shift start - FE	X	X	X	X
Acuity code - FE	X	X	X	X
Chief complaint - FE	X	X	X	X
Physician - FE		X		X
Mean of DV	4.55	4.55	0.24	0.24
N	76,001	76,001	76,001	76,001
R^2	0.204	0.246	0.221	0.243
Adjusted R^2	0.196	0.232	0.213	0.229

Notes: The table presents OLS estimates of linear regressions of the number of diagnostic orders and a dummy variable whether the patient was admitted to the hospital during the encounter on decile indicators of mean of the leave-out measure of cognitive load during the encounter (standardized to mean 0 and SD 1). The unit of observation is the *encounter*. All specifications control for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, and acuity code. Columns (2), (4), and (6) additionally include physician fixed effects. Robust standard errors are shown in parentheses. Statistical significance is indicated by *** (1%), ** (5%), and * (10%).

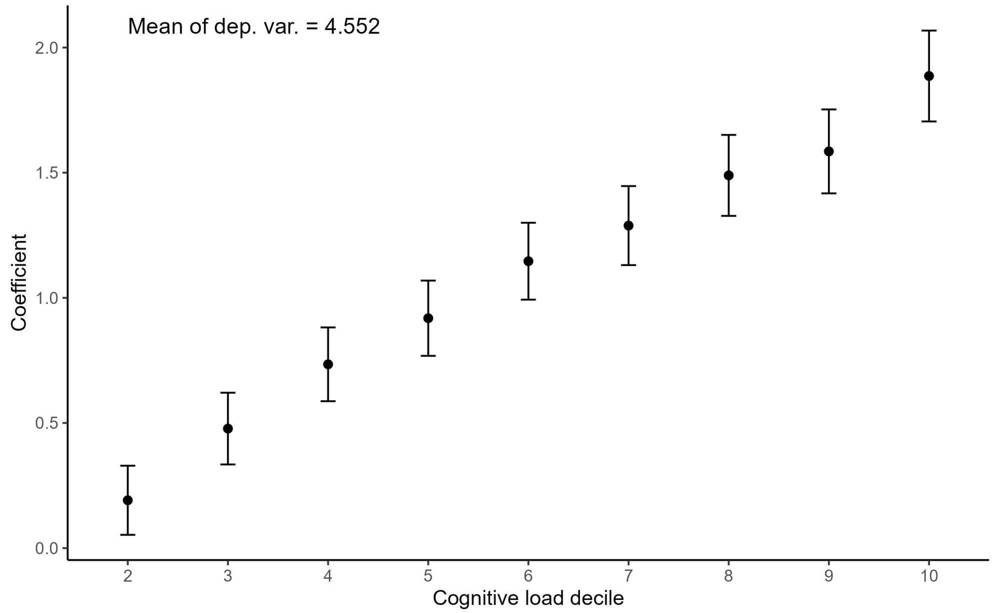


Figure G.1: Estimates of the Impact of Decile of Cognitive Load on Number of DX Orders

Notes: The figure plots coefficients from linear regressions of the number of diagnostic orders on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter. The unit of observation is the encounter. The specification controls for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, acuity code, and physician fixed effects. The first decile is the omitted category. Estimates correspond to Column (2) of Appendix Table G.4. Robust standard errors are used.

Table G.10: Patient-share Transition Matrix from Actual to Counterfactual Physician Category

	Counterfactual physician type				
	Resident Years 1/2	Resident Years 3+	Attending physician	Nurse practitioner	*Sum*
Resident Years 1/2	0.226	0.062	0.077	0.055	0.419
Resident Years 3+	0.081	0.097	0.048	0.036	0.262
Attending physician	0.047	0.022	0.028	0.019	0.145
Nurse practitioner	0.062	0.029	0.031	0.052	0.174
Sum	0.416	0.210	0.184	0.162	1.000

Notes: Each row reports the share of encounters that move from the physician category listed in the stub (“Actual”) to the category in the corresponding column (“Counterfactual”). The final column shows the row sum; the final row shows column sums, and the entire table sums to 1.

G.3 Further health outcomes

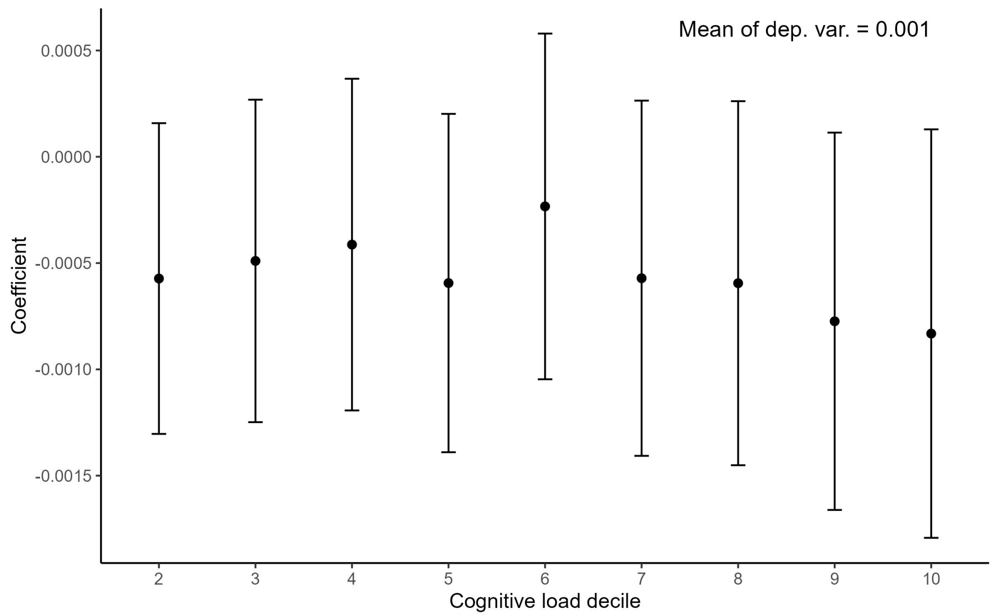


Figure G.2: OLS Estimates of the Impact of Decile of Cognitive Load on ED Death

Notes: The figure plots coefficients from linear regressions of an indicator for whether the patient died during the ED stay on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter. The unit of observation is the encounter. The specification controls for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, acuity code, and physician fixed effects. The first decile is the omitted category. Robust standard errors are used.

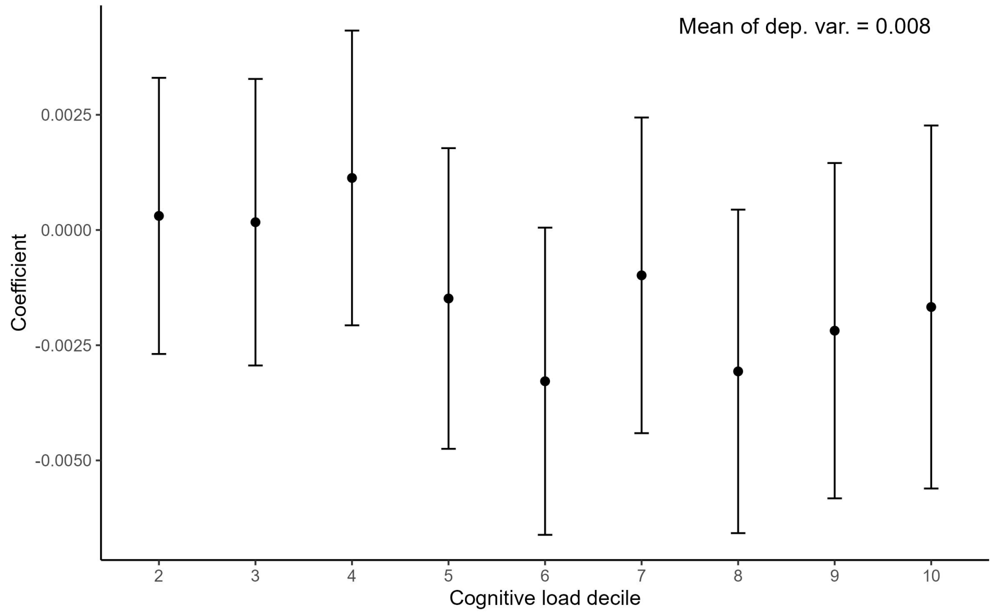


Figure G.3: OLS Estimates of the Impact of Decile of Cognitive Load on Inpatient Death

Notes: The figure plots coefficients from linear regressions of an indicator for whether the patient died during the inpatient stay on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter. The unit of observation is the encounter. The specification controls for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, acuity code, and physician fixed effects. The first decile is the omitted category. Robust standard errors are used.

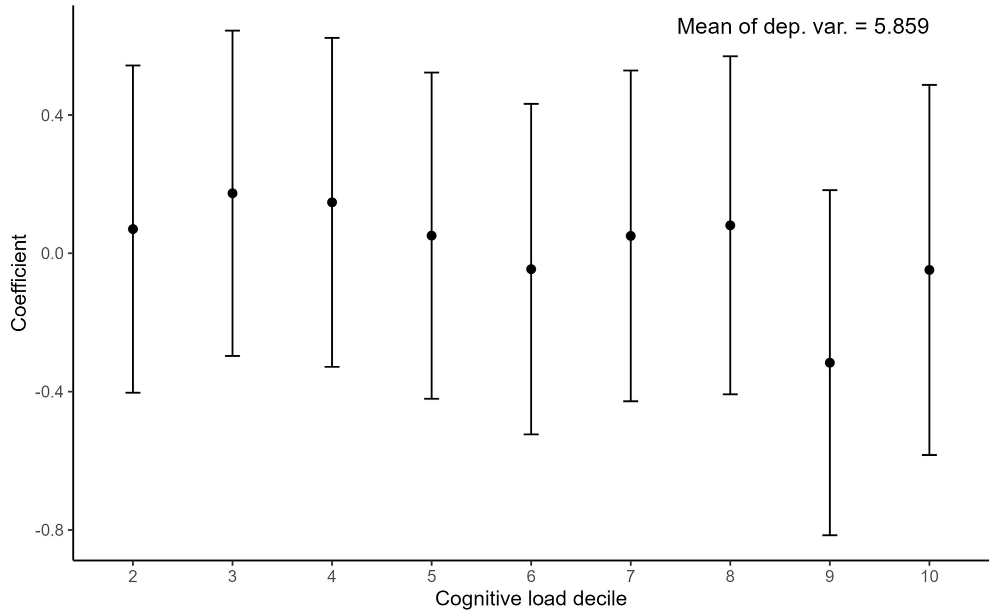


Figure G.4: OLS Estimates of the Impact of Decile of Cognitive Load on Inpatient Length of Stay

Notes: The figure plots coefficients from linear regressions of inpatient length of stay (days) on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter. The unit of observation is the encounter. The specification controls for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, acuity code, and physician fixed effects. The first decile is the omitted category. Robust standard errors are used.

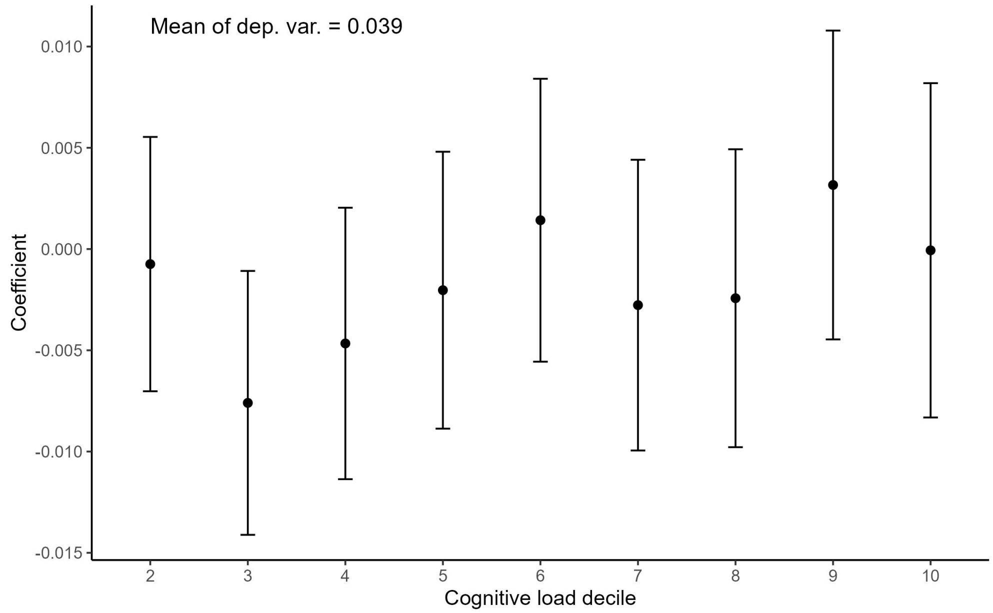


Figure G.5: OLS Estimates of the Impact of Decile of Cognitive Load on 30-Day Readmission

Notes: The figure plots coefficients from linear regressions of an indicator for whether the patient was readmitted within 30 days on decile indicators of the maximum of the leave-out measure of cognitive load during the encounter. The unit of observation is the encounter. The specification controls for max prior cognitive load and include fixed effects for the maximum shift hour during which the physician interacted with the encounter, chief complaint, acuity code, and physician fixed effects. The first decile is the omitted category. Robust standard errors are used.

H Physician Heterogeneity in Cognitive Load Effects

The objective of this Appendix section is to move beyond average effects and ask whether the impact of cognitive load differs systematically across provider characteristics such as provider category (attending physician vs. nurse practitioner vs. resident), gender, tenure, specialty, and (in a smaller sample) medical school ranking and year of graduation.

We use the *encounter-action*-level panel, with our standard explanatory variables as the standardized measure of contemporaneous cognitive load at the time of the action. As in our main specifications, we residualize both the outcome and treatment against prior cognitive load, and fixed effects for hour-of-day, day-of-week, hours-since-shift-start, hours-until-shift-end, provider-identifier, acuity-code, and chief-complaint.

We begin with the larger subsample for which we observe provider characteristics supplied directly by UCSF. In this subsample, the available heterogeneity variables are provider role, gender, tenure, and specialty setting. Provider role is captured through indicators for nurse practitioners or physician assistants (NP/PA), residents in postgraduate years 1-2, and residents in postgraduate years 3+, with attending physicians as the omitted category. Gender is measured with an indicator for male providers, so that female providers form the omitted category. Provider experience within the institution is summarized using tenure quartiles, entered through indicators for the second, third, and fourth quartiles relative to the first quartile. Finally, specialty setting is captured by an indicator for providers who are not in emergency medicine, with emergency medicine as the omitted category. These UCSF-provided characteristics are available for 482 out of 610 providers, and this subset accounts for 185,058 out of 266,625 order events in the encounter-action sample.

We then consider a smaller linked subsample in which providers can be matched to additional background characteristics using their National Provider Identifier (NPI). This linked sample contains 208 out of 610 providers and covers 113,592 out of 266,625 order events in the encounter-action data. In this subsample, we retain the same core provider characteristics as above – provider role, gender, tenure, and specialty setting – and add two further dimensions of heterogeneity: medical school ranking and training cohort. Medical school quality is measured using indicator variables for rank groups 1-5, 6-10, 11-15, 16-20, and 21-25, with an additional "other" category for ranks above 25 serving as the omitted group. Training cohort is summarized using graduation-decade indicators for the 1970s, 1980s, 1990s, and 2000s, with the latter serving as the omitted category. This two-sample structure is useful because it allows us to study a broad set of provider traits in the larger UCSF-characteristics sample, while also examining whether richer background characteristics available only through NPI linkage provide additional explanatory power in the smaller linked sample.

H.1 Causal Forest Estimation and Summary Objects

Let Y_i denote our outcome, W_i denote the measure of contemporaneous cognitive load, and X_i denote the vector of provider characteristics. For ease of exposition, we only use the *i*-subscript. For each outcome and sample, we fit a causal forest using the residualized variables $(\tilde{Y}_i, \tilde{W}_i)$ and moderators X_i to estimate heterogeneous effects (Wager and Athey, 2018; Athey et al., 2019):

$$\tilde{Y}_i = \tau(X_i)\tilde{W}_i + \varepsilon_i,$$

where $\tau(\cdot)$ is an unknown function capturing heterogeneity in the marginal effect of cognitive load intensity.

Causal forests are an extension of random forests designed to estimate heterogeneous treatment effects rather than conditional means. The object of interest is the conditional (marginal) effect function $\tau(x)$ – in our setting, how a one-standard-deviation increase in residualized cognitive load $\tilde{W}_i$ changes a residualized outcome $\tilde{Y}_i$ as a function of moderators X_i. The algorithm builds an ensemble of trees that recursively partition the moderator space to form subgroups with systematically different estimated effects. Crucially, each tree is grown using "honesty," meaning that the data used to choose the splits (tree structure) are kept separate from the data used to estimate within-leaf effects: this separation reduces adaptive overfitting and yields more reliable out-of-sample treatment-effect predictions. We implement causal forests using the `grf` R package following Wager and Athey (2018) and Athey et al. (2019). Concretely, for each tree we draw a 50% subsample of the full dataset ("in-bag") and retain the remaining 50% as out-of-bag observations for out-of-sample prediction and inference. With honesty, the in-bag half is split evenly again: 25% of the full sample is used to learn the tree structure (splits), and a disjoint 25% is used to estimate within-leaf treatment effects conditional on that structure. Each tree therefore yields a local estimate of $\tau(x)$ by comparing outcomes across observations with different $\tilde{W}_i$ within the same leaf, and the forest aggregates these local estimates across many trees using adaptive nearest-neighbor weights to produce $\hat{\tau}(X_i)$ for each observation. We use standard `grf` implementation defaults (`num.trees`=2000, `sample.fraction`=0.5, `honesty`=TRUE with `honesty.fraction`=0.5, and a small `min.node.size`). Because the unit of observation in our estimation is the encounter-action, all ATE and heterogeneity summaries are action-weighted: providers who generate more actions contribute more to the objective and to aggregated estimates.

Average effect (ATE): For each outcome and sample, we report the average treatment effect implied by the forest,

$$\widehat{\text{ATE}} \equiv \frac{1}{N}\sum_{i=1}^{N}\hat{\tau}(X_i),$$

with standard errors from the forest's inference procedure using encounter clustering. The average treatment effect (ATE) is the average of the conditional average treatment effects (CATEs) over the estimation sample.

Best linear projection (BLP): To provide an interpretable linear summary of heterogeneity, we use the best linear projection routine and report coefficients from the projection of $\widehat{\tau}(X_i)$ on X_i:

$$\widehat{\tau}(X_i) = \beta_0 + X_i'\beta + u_i.$$

When X_i consists of indicator variables, each element of β is interpretable as a partial association between the indicator and the estimated conditional effect, holding other moderators fixed.

Figure H.1 shows the estimated best linear projection (BLP) coefficients by outcome and sample. Each point corresponds to the coefficient on a provider characteristic in the linear projection of the forest-predicted conditional treatment effects on the moderator vector, and the whiskers denote 95% confidence intervals. The top panel reports results for the larger UCSF-characteristics sample, in which the available moderators are provider role, gender, tenure, and specialty setting. The bottom panel reports results for the smaller NPI-linked sample, which includes the same core moderators and additionally incorporates medical school rank bins and graduation-decade indicators. The figure is therefore best interpreted as a compact summary of which observable provider characteristics are most strongly associated with heterogeneity in the marginal effect of contemporaneous cognitive load across outcomes.

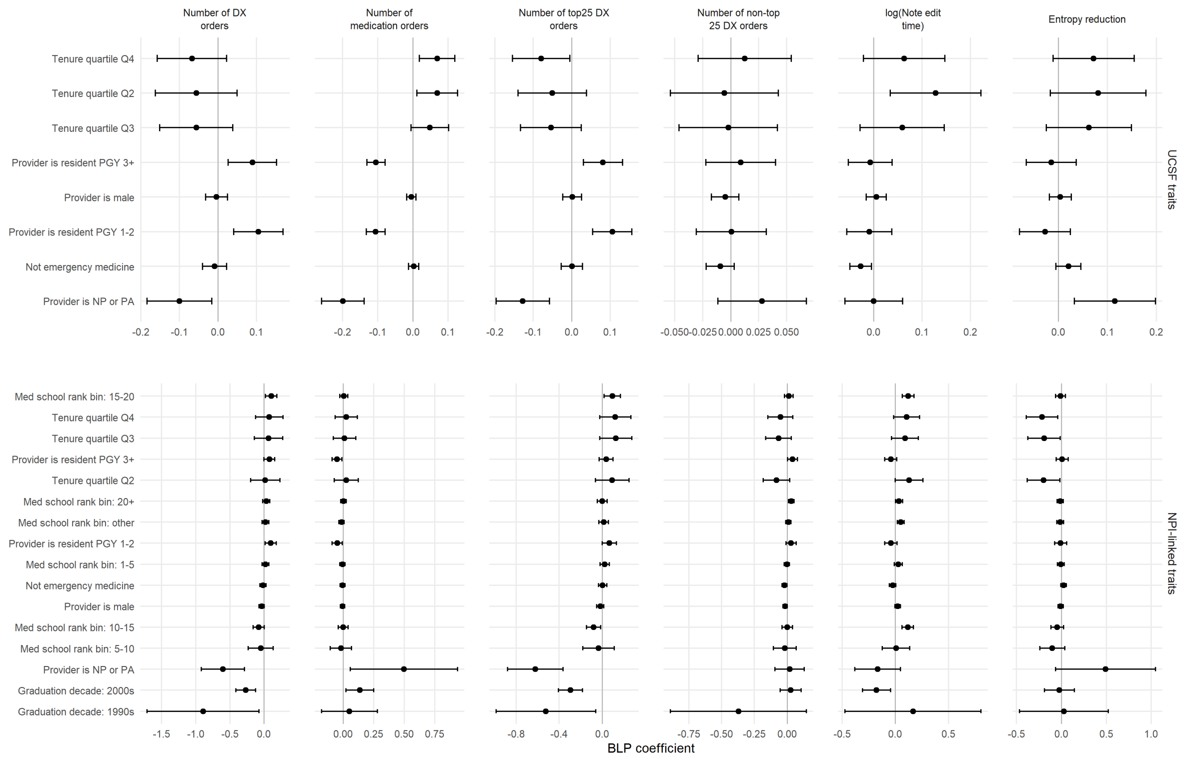


Figure H.1: Best Linear Projection (BLP) Coefficients by Outcome and Sample

Notes: The figure plots best linear projection (BLP) coefficients summarizing heterogeneity in the marginal effect of contemporaneous cognitive load across provider characteristics. For each outcome (columns), we estimate a causal forest on the encounter-action panel using residualized outcome and treatment variables and provider characteristics as moderators. Points report BLP coefficients from a projection of estimated conditional average treatment effects (CATEs) on the moderator indicators; whiskers denote 95% confidence intervals based on encounter-clustered standard errors. The top panel uses UCSF-provided provider traits (role, gender, tenure quartiles, and specialty setting). The bottom panel uses the NPI-linked sample and additionally includes medical-school-rank bins and graduation-decade indicators. Outcomes and cognitive load are residualized on prior cognitive load and fixed effects for hour-of-day, day-of-week, hours-since-shift-start, hours-until-shift-end, acuity code, chief complaint, and provider fixed effects. Effects are per 1 SD increase in cognitive load.